

The University of Chicago

AN EXPERIMENTAL STUDY OF
POTENTIAL FAILURES IN COLLEGE

A PART OF A DISSERTATION SUBMITTED
TO THE FACULTY OF THE DIVISION OF THE
SOCIAL SCIENCES IN CANDIDACY FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATION

1934

By
STEPHEN PAUL RYDER

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

1934

The University of Chicago

AN EXPERIMENTAL STUDY OF
POTENTIAL FAILURES IN COLLEGE.

A PART OF A DISSERTATION SUBMITTED
TO THE FACULTY OF THE DIVISION OF THE
SOCIAL SCIENCES IN CANDIDACY FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATION

1934

By

STEPHEN PAUL RYDER

Private Edition, Distributed by
THE UNIVERSITY OF CHICAGO LIBRARIES
CHICAGO, ILLINOIS

1934



TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. THE GENERAL PROCEDURE AND METHODOLOGY OF THE STUDY .	8
III. THE CONSTRUCTION AND PRELIMINARY USE OF THE DIAGNOSTIC MATERIALS	22
IV. THE ORGANIZATION OF THE DIAGNOSTIC AND REMEDIAL EXPERIMENT	59
V. THE EXPERIMENTAL PROGRAM OF DIAGNOSIS AND REMEDICATION	74
VI. FINDINGS AND INTERPRETATIONS OF THE STUDY	100

ACKNOWLEDGMENTS

The writer is deeply indebted to the following persons: to William S. Gray for counsel in identifying the problems and for advice at all stages of the investigation; to Guy T. Buswell, Karl J. Holzinger, Frank N. Freeman, Rollo S. Lyman, and Aaron J. Brumbaugh for critical reading of the manuscript; to William Reitz for permission to apply his prediction formulas; to innumerable investigators in reading and study who offered constructive diagnostic and remedial suggestions in their published reports; and to his wife, Ruby Louise Ryder, who rendered invaluable assistance throughout this study.

CHAPTER I INTRODUCTION

The Purpose of the Study

Recent research studies have shown that a considerable portion of our college population has failed to acquire, in its elementary and secondary-school training, the fundamental skills necessary for efficient silent reading.¹ Other studies have shown that many college students lack that knowledge of and use of good study procedures which are essential to effective and economical learning.² The importance of these two findings lies in the fact that a significant positive relationship exists between actual scholastic success and reading³ and study⁴ abilities.

In consideration of these facts, higher institutions have attempted to adjust the individual student to the institution through the use of diagnostic and remedial procedures in reading and study.⁵ Studies made in this connection indicate, however, that the diagnostic and remedial techniques in these two funda-

¹Floyd W. Reeves and Others, Instructional Problems in the University, p. 148. University of Chicago Survey, Vol. IV. Chicago: The University of Chicago Press, 1933.

²William F. Book, Learning How to Study and Work Effectively, p. 17. Boston: Ginn and Company, 1927.

³Elbert K. Fretwell, A Study of Educational Prognosis. Teachers College Contributions to Education, No. 99. New York: Teachers College, Columbia University, 1919.

⁴Luella C. Pressey, "What Are the Crucial Differences between Good and Poor Students," Research Adventures in University Teaching, pp. 4-10. Bloomington, Illinois: Public School Publishing Company, 1927.

⁵William S. Gray, "Provision for the Individual in College Education," Vol. IV. Proceedings of the Institute for Administrative Officers of Higher Institutions, 1932. Chicago: The University of Chicago Press, 1932.

mental academic abilities are inadequate, at present, at the college level.¹

The purpose of the present study arises naturally from the facts that have been presented. It comprises two major aspects: (1) to construct and partially standardize a comprehensive diagnostic instrument in reading and study at the college level, which makes possible the early differentiation between potentially superior and inferior students for purposes of subsequent remediation; and (2) to apply the comprehensive diagnostic instrument and related remedial procedures to potential college failures, under experimental conditions, to measure their administrative value.

Definition of Terms

The term reading, as used throughout the study, may be defined as "the act, practise, or art of perusing written or printed matter and ascertaining or considering its contents and meaning."² This comprehensive definition of reading is used because the diagnostic technique developed in the study includes all of the major types of reading activities, especially "analytical study" which has rightly been differentiated from the ordinary reading process.³

A similarly comprehensive definition is used for the term study, namely, "the act of applying thought to a subject; hence the application of the mind to books, to art or science, or to any subject of investigation."⁴ This broad definition is adopted for essentially the same reason that was given for the use of the definition of reading.

The term diagnosis may be defined as "the procedure used to identify or to determine the nature of a deficiency or weak-

¹Ibid., pp. 144-58.

²Funk and Wagnalls, New Standard Dictionary of the English Language. New York: Funk and Wagnalls Co., 1924.

³Charles H. Judd and Guy T. Buswell, Silent Reading: A Study of the Various Types, pp. 47-89. Supplementary Educational Monographs, No. 23. Chicago: Department of Education, University of Chicago, 1922.

⁴Funk and Wagnalls, op. cit.

ness.¹ The diagnostic instrument developed in the study is qualified by the adjective comprehensive owing to the extreme diversity of the materials of which it is composed and to the completeness of the diagnosis which may be obtained through its use.

No satisfactory definition could be found in the literature for the term remediation. As used in the present study, however, it may be defined simply as the application of corrective work designed to overcome the existing faults or deficiencies revealed by the diagnosis.

Need for the Study

A critical analysis of previous investigations in reading and study, the most important of which are listed in the final bibliography, revealed the fact that there is a definite need, at the college level, for a comprehensive study of the proposed type. Such evidence is summarized at length by Gray² and Cowley³ and will not be reviewed here.

In addition to the general need for the proposed study, as reported in previous investigations, there are three specific reasons for such a study at the University of Chicago. The first of these is found in the radical changes⁴ involved in the organization and administration of the "New Plan" of that institution. Freedom from the compulsory aspects of class attendance, quarterly examinations, and other traditional or conventional college regulations places larger responsibilities upon the student for independent work. This is true also in many other types of experi-

¹Leo J. Brueckner and Ernest O. Melby, Diagnostic and Remedial Teaching, p. 123. New York: Houghton Mifflin Co., 1931.

²William S. Gray, Summary of Investigations Relating to Reading. Supplementary Educational Monographs, No. 28. Chicago: Department of Education, University of Chicago, 1925. Also Supplements 1-5.

³W. H. Cowley, The Personnel Bibliographical Index. Bureau of Educational Research, Ohio State University, 1932.

⁴Reeves and Others, op. cit., IV, 112-16.

mentation¹ at the college level, including "Honors Courses," and the "Tutorial" procedure. The enjoyment of the fullest fruits of the opportunities provided, necessitates a knowledge on the part of the student of how to use his time or freedom most effectively from an academic viewpoint. Among other things, he should know and use the best possible procedures in two fundamental types of academic activity, reading and study.

Furthermore, the comprehensive examinations used at the University to measure objectively and impartially the student's knowledge in the four general survey courses in the Social Sciences, the Biological Sciences, the Physical Sciences, and the Humanities cover a wide range of textual and supplementary reading materials. The preparation and organization of this material places a premium upon efficient reading and study procedures. The measures developed in this investigation for the diagnosis and remediation of these procedures have especial significance, therefore, under "New Plan" conditions.

An additional reason for the study at the University of Chicago is found in facts relating to the counseling load. Each member of the staff of advisers in the Office of the Dean of Students has approximately 85 freshmen and 85 sophomores, or a total of 170 students, under his direction. In addition, the advisory staff has a considerable load of instructional duties which makes the effective personal counseling of this large group of students exceedingly difficult without some adequate and specific diagnostic instrument. Similarly, the instruction and guidance given in connection with English 102, which deals with "Note-taking," "Instruction in the preparation of original papers," and "Training in handling the commoner types of written assignments made in University courses" do not seem to have reduced the percentage of academic failure to any considerable extent. Ninety-seven per cent of the 50 failing students considered in this study had registered in English 102 during the academic year in which they failed.

¹"Changes and Experiments in Liberal Arts Education," The Thirty-First Yearbook of the National Society for the Study of Education, Part II, p. 5. Bloomington, Illinois: Public School Publishing Company, 1932.

The problem of student retention at the University of Chicago also differs somewhat from that in many other institutions. In summarizing the results of investigations dealing with student mortality in 132 colleges and universities of all types, May¹ has shown that there is an annual freshman mortality, in the group of higher institutions studied, ranging from 23 to 37 per cent. The total mortality figures include dismissals for poor scholastic work in addition to other non-scholastic causes of dismissal or withdrawal. These data are representative, in part, of the retention problem existing at the University of Chicago. Table I presents mortality data for three "Old Plan" and two "New Plan" classes, in terms of the students dismissed for poor work, students withdrawn while on probation, and students on probation at the end of the first year.

TABLE I
PERCENTAGE OF FRESHMEN WHO MADE UNSATISFACTORY
SCHOLASTIC RECORDS DURING THEIR FIRST YEAR
IN THE UNIVERSITY OF CHICAGO²

Status of Students at the Close of the First Year	1926-27	1927-28	1928-29	1931-32	1932-33
Dismissed for poor work	8.4	4.1	4.4	5.6	7.6
Withdrew while on probation . . .	3.6	10.0	6.4		
On probation at end of first year .	16.7	15.3	12.7	15.9*	14.3*
Total unsatis- factory . . .	28.7	29.4	23.5	21.5	21.9

*Data are not available to show the probationary status for these years since there is no probation under the "New Plan." The percentage of students who have not met any of their seven college requirements at the beginning of their second year in college may be taken, however, as evidence of unsatisfactory progress corresponding approximately to probationary status under the "Old Plan" (The University of Chicago Handbook for Students, p. 12. Chicago: The University of Chicago Press, May, 1934. See Section

¹E. O. May, "The Holding Power of the New College Plan of the University of Chicago." Unpublished report to the Dean of Students in the College. University of Chicago, January, 1933.

²The data for the three "Old Plan" classes, 1926-29, are adapted from the University of Chicago Survey, V, 22.

H, "Minimum Requirements for Continuance in College.>"). Such percentages are substituted therefore in Table I for the "New Plan" classes. The term "student dismissal" is used under the "New Plan" when a student makes no mark but F in all examinations taken at the end of the first academic year, or when he fails to pass at least one-half of the comprehensive examinations for which he is presumed to be prepared by the courses which he has pursued.

The data presented in this table show clearly the general retention problem which exists at the University of Chicago under "New Plan" conditions. More than one-fifth of the freshmen make unsatisfactory scholastic records during their first year in the University. The importance of the problem is emphasized further by the fact that the students actually dismissed for poor work did not receive a major portion of the total number of unsatisfactory marks earned on the comprehensive examinations. The number and percentage of F or failing marks earned by two "New Plan" classes, by students dismissed for poor work, and by students not dismissed for poor work are shown in Table II.

TABLE II
NUMBER AND PERCENTAGE OF F OR FAILING MARKS RECEIVED
BY TWO "NEW PLAN" CLASSES, BY STUDENTS DISMISSED
FOR POOR WORK, AND BY STUDENTS NOT DISMISSED
FOR POOR WORK

Groups	F Marks Received by Total Class		F Marks Received by Students Dismissed		F Marks Received by Students Not Dismissed	
	No.	Per Cent	No.	Per Cent	No.	Per Cent
Class of 1931-32	223	100	79	35.4	144	64.6
Class of 1932-33	186	100	85	45.7	101	54.3
Combined Classes	409	100	164	40.1	245	59.9

An analysis of the data presented in this table indicates that the percentage of F or failing marks received by the students actually dismissed for poor academic work is less than half of the total percentage of F marks earned on the examinations. A consideration of this fact leads to the conclusion that students who are not potential college failures in all courses are in need,

nevertheless, of some remedial treatment, possibly of a more specific type than that which is needed by students who have a general deficiency.

There are three major reasons therefore for the specific need of the study at the University of Chicago. The student freedom and the comprehensive examinations of the "New Plan" necessitate the utmost efficiency in reading and study on the part of individual students. The counseling and instructional loads of the college advisory staff are so heavy that effective personal guidance cannot be accomplished without an adequate and specific diagnostic instrument. Finally, the serious retention problem which exists at the University indicates the need for adequate measures for the diagnosis and remediation of reading and study deficiencies as a possible means of redeeming potential college failures.

Organization of the Report

Following Chapter II in which the methodology of the study is discussed, the organization of the report conforms to the two-fold purpose of the study that was indicated in the opening section of the chapter. The construction and preliminary use of the diagnostic materials are considered in Chapter III; and the organization, administration, and results of the diagnostic and remedial experiment are considered in Chapters IV and V. In Chapter VI the general findings and contributions of the study are presented, interpretations are made, and suggestions are given for the further study and use of the diagnostic instrument and related remedial procedures by college administrators.

CHAPTER II

THE GENERAL PROCEDURE AND METHODOLOGY OF THE STUDY

Problem of the Chapter

The procedure followed in the construction, partial standardization, and experimental use of the diagnostic instrument is more or less complex and differs to a considerable degree from those reported in previous studies. In order that a complete and integrated over-view of the investigation may be provided before presenting the detailed description of the diagnostic materials and their use, a preliminary outline of the steps of procedure is included in this chapter. Furthermore, the partial standardization of the diagnostic instrument necessitated the selection of statistically adequate student samples. In order to simplify the presentation of the diagnostic materials in later chapters, the statistical and technical considerations involved in the selection of these samples and in the general treatment of the data are discussed also in this chapter.

Outline of the General Procedure

The two major steps of procedure used in the study, which correspond to the twofold statement of purpose presented in the preceding chapter, are indicated by the following statements:

(1) The comprehensive diagnostic instrument was constructed, administered, and partially standardized.

(2) The diagnostic instrument and related remedial procedures were applied to potential college failures under experimental conditions designed to measure their administrative value.

Each of these major steps or divisions of procedure included, however, a series of detailed sequential steps which are summarized in modified outline form:

- I. The construction, administration and partial standardization of the diagnostic instrument involved the following steps:
 - A. A critical survey and analysis was made of all available literature on reading and study to determine the items or

factors which were of considerable diagnostic and remedial significance.

- B. As a result of the critical analysis of the available literature, a comprehensive diagnostic instrument was developed to measure the most significant factors or aspects of reading and study. The instrument consisted of a battery of five placement tests, 21 informal individual tests, 19 reading and study check lists, personal rating scales, and exploratory personal and family data sheets.
- C. The raw scores earned on the tests and sub-tests of the placement battery by typical samples of 100 students were obtained from the Board of Examinations. These samples of 100 were carefully selected to give a relatively perfect cross-section of the 1932-33 class of 700 students in each ability or test.
- D. The remainder of the diagnostic materials, with the exception of the data sheets, was administered to one typical sample of 100 students. The latter sample was carefully selected to give a relatively perfect cross-section of the performance of the same class of 700 students on the comprehensive examinations in the four general college courses. Facilities were not available for the administration of this portion of the diagnostic technique to the total group of 700. This fact necessitated the selection of the typical sample of 100.
- E. The scores obtained from the administration of the diagnostic materials to these samples of 100 were used in the construction of a comprehensive five-point rating scale upon which the raw scores of any student might be projected, subsequently, and given the proper rating. The use of this rating scale device enabled the experimenter to derive desired summations or combinations of scores on all diagnostic materials on a comparable basis without resorting to the use of standard scores.
- F. The diagnostic materials were administered to 50 superior, 50 middle, and 50 inferior students of the total class of 700 who entered the University of Chicago in September, 1932. Performance on the comprehensive examinations, June 1933, was taken as the criterion of selection. The 50 superior students were those who received either three

A's or two A's and one B. The 50 middle students were those who received at least two C's and no mark other than C. The inferior students were those who received at least two F's.

- G. The raw scores obtained from the administration of the diagnostic materials to the 50 superior, 50 middle, and 50 inferior students were projected upon the comprehensive rating scale to obtain summations and combinations of scores on a comparable basis. From the ratings obtained in this manner, a second group of norms was derived that was composed of the mean scores of the three groups, superior, middle, and inferior. Through the use of these norms a three-point classificatory scale was constructed upon which the scores of any student on the diagnostic materials might be projected, subsequently, for purposes of classification or differentiation into the proper categories. The middle group was used throughout the study to delimit the central area of the scale and to clarify the relative areas occupied by the superior and inferior students.

- H. From the responses obtained by the administration of the exploratory personal and family data sheets to the 50 superior and 50 inferior students, further conclusions were drawn concerning the relationship existing between scholastic efficiency, reading and study abilities, and certain other factors. These included age, sex, intelligence, social and economic background, personality, previous academic training and marks, patterns of courses taken, vocational and avocational interests, size and ancestry of families, amount of reading done, size of home and school libraries, club membership, type of extra-curricular activities engaged in, and many others. An attempt was made thus to determine or explore those factors or areas of experience, training, or environment which contribute, genetically, to the development of good and poor students.

- II. The diagnostic instrument and related remedial procedures were applied to potential college failures under experimental conditions to measure their administrative value:

- A. A statistical investigation was made of the two previous classes, 1931-32 and 1932-33, entering the University of

Chicago under the "New Plan." Effective and reliable means of identifying potential college failures were determined from a study of the relationship existing between marks on the comprehensive examinations and such factors as average of high-school marks, rank in high-school graduating class, and percentile ranks on the placement tests in intelligence and reading comprehension. (This preliminary identification of potential college failures made it unnecessary to administer the complete diagnostic instrument to the entire class of entering freshmen, September 1933. The instrument could thus be administered to those students who were classifiable, in general, as potential college failures and the degree of deficiency in reading and study would indicate whether or not remedial procedures should be instituted as a possible means of student retention.)

- B. These preliminary predictive measures were applied to the class of 700 freshmen entering the University in the autumn quarter, 1933-34. A group of 39 potential college failures was identified. Since the matched pair and parallel group methods of experimentation were to be used, a preliminary matching of these students on the college courses in which they were registered was made to eliminate those who could not be equated exactly on this essential variable. It was found that only 26 of the 39 potential college failures could be equated exactly on this factor.
- C. The battery of diagnostic materials was administered to this group of 26 potential college failures to obtain a preliminary diagnosis of deficiency. Three of the 26 subjects were eliminated from the experiment, since they did not have a marked degree of deficiency in reading and study, thus reducing the total number of experimental subjects to 23.
- D. This group of 23 potential college failures, who had considerable deficiency in reading and study, was divided into pairs of students who were matched discriminately on all of the essential variables which the investigator desired to hold constant during the operation of the experimental program of remediation.

- E. On the basis of the diagnosis and further analysis of the specific items of deficiency, relatively intense, individual, or clinical remedial procedures in reading and study were used with the 10 students finally selected for Experimental Group A. The 19 reading and study check lists used in this connection were composed of 931 specific items recommended by specialists as good procedures. These lists furnished the basis for a diagnosis of the specific items of deficiency as well as a series of remedial guides to the best procedures in reading and study. They were used as the chief basis for the program of remediation, supplemented by vocabulary lists in the four general survey courses of the college. The students comprising Experimental Group B were given the vocabulary lists and a complete set of the reading and study check lists to be used as remedial guides for self-directed-study. The control group was used only for comparative purposes in the initial and final testing programs, and to hold constant the essential variables of the experiment. The remedial program was carried forward for a period of 29 weeks. One hour each week was devoted to the remedial guidance of each subject in Experimental Group A.
- F. A final testing program was administered to the total group of 23 potential college failures at the close of the remedial experiment and the results achieved on the college comprehensive examinations by these subjects were obtained. The effectiveness of the diagnostic instrument and of the related remedial procedures was measured and evaluated, finally, in terms of improved scholastic success and in terms of increased personal efficiency in reading and study.

The preceding outline of the general procedure summarizes the scope of the investigation and the procedure followed in achieving the twofold purpose of the study. A more detailed report of each procedure is presented in succeeding chapters where the construction, partial standardization, and use of the diagnostic materials are discussed, and the diagnostic and remedial experiment is reported.

Sampling Methods

The selection of statistically adequate and reliable student samples presented a problem of major importance in connection with the development of comparative norms for the diagnostic instrument. The complete battery of diagnostic materials was too lengthy to be administered by the investigator, individually, to the total class of 700 students, 1932-33. The selection of adequate samples, however, rendered the use of the entire class unnecessary. The general statistical considerations involved in the sampling procedure are discussed at this point to simplify the later presentation of the diagnostic materials.

The typical samples of 100.--After the preliminary development of the diagnostic materials, one of the first problems encountered in the study was the construction of a statistical device by means of which the raw scores on all of the diverse diagnostic measures could be reduced to some comparable, simple, and readily understandable basis. A complex system of standard scores, which required considerable computation, would militate against the actual use of the diagnostic instrument by college administrators or counselors. This would defeat the practical administrative purpose for which the instrument was designed. To meet this need a relatively simple rating scale device was constructed, composed of the raw scores of selected samples of the entire class of 700 students, upon which all subsequent scores of students could be projected and given a comparable five-point rating. The chief sample used in the development of comparative norms has been termed therefore a typical sample. It was not a random sample inasmuch as it was carefully selected to represent typically the parent population of 700 students.

Thirty-four of these typical samples were used in the construction of the comprehensive rating scale device, including 100 individual students or student scores in each sample. They were selected in such a way that a complete cross-section of the total class was secured on each ability or test, since the sample which represented the total population in intelligence might not represent it, necessarily, in reading rate, personality, or high-school average. These samples were therefore distinctly different but not necessarily mutually exclusive on each of the placement tests and sub-tests administered by the Board of Examinations. In the administration of the diagnostic materials constructed by

the investigator, however, only one typical sample of 100 students was used. It was selected on the basis of the performance of the students on the college comprehensive examinations in the four general survey courses.

The sampling procedure used in the case of the placement tests administered by the Board of Examinations of the University of Chicago was as follows: (1) raw score distributions of the total class of 1932-33 were obtained from the Board of Examinations for each test and sub-test; (2) the typical sample of 100 was obtained from the parent distribution by entering each frequency interval with a statistical constant derived by dividing the number desired in the sample population by the total number in the parent population; (3) each frequency interval in the parent distribution was made to yield its proportionate share to the sample population by multiplying the number in the parent frequencies by the statistical constant derived previously; and (4) when the necessary number of frequencies was obtained in the sample population, individual scores were withdrawn at random from within each frequency interval of the parent population to obtain the desired raw scores for the sample.

The number of individual scores to be withdrawn from each frequency interval was indicated by the product obtained by multiplying the number of frequencies in each interval of the parent population by the proper statistical constant. By the use of this method 100 individual scores were obtained which gave a typical sample of the parent population on each test and sub-test. These 100 scores were then ranked in ascending order of ability to form the basis upon which a five-point rating scale device was constructed by selecting the 1st, 20th, 40th, 60th, 80th, and 100th scores and giving the appropriate rank to the five intervals thus created.

In a similar manner, a parent distribution composed of all marks and all combinations of marks earned by the class of 1932-33 on the comprehensive examinations, June 1933, was taken as the basis for the selection of the 100 individual students used in the administration and partial standardization of the diagnostic materials constructed by the investigator. The latter sample was selected on this basis for the reason that the cross-section of general abilities which were to be measured in this phase of the study would, doubtless, be present in such a sample

in greater degree than in samples chosen on the basis of any one measure such as intelligence, reading rate, or reading comprehension.

The parent and sample distributions used in the selection of the latter sample of 100 are shown in Table III to illustrate the general method of selection, which is identical for all samples of 100 used in the study.

TABLE III
DISTRIBUTIONS OF PARENT AND SAMPLE POPULATIONS ON THE
BASIS OF MARKS EARNED ON THE COMPREHENSIVE
EXAMINATIONS, JUNE, 1933

Combinations of Marks		Number in Parent Population	Number in Sample* Population
All	C	106	18
	CD	95	16
	BC	94	16
	FC	44	7
	AB	41	7
All	F	31	5
	FD	29	5
All	B	28	5
All	A	26	4
	CDF	26	4
	AC	22	4
	ABC	18	3
All	D	15	2
	BCD	9	1
	BCF, BF, ACD . .	6	1
	BD	5	1
	BDF, ACF, ACD, FB	5	1
Total		600	100

*The constant used to enter each frequency interval of the parent distribution was $100/600$, or .1666.

Since each frequency interval of the parent distribution in Table III was multiplied by a statistical constant which made the interval yield its proportionate share of frequencies to the sample population, the frequency distribution of the sample was representative of the parent distribution. Each of the frequencies

within an interval corresponds to a definite raw score. The desired raw scores in the sample population could be obtained therefore by simply entering each interval in the parent distribution and selecting at random the proper number of scores made by individual students. In selecting the 100 individual students in the sample shown in Table III, however, the random procedure was not followed. In order to make this sample as representative as was statistically possible, individual students were selected from each interval of the parent distribution who would contribute the proper combinations of marks that would give an exact duplication of the percentage distribution of marks earned on the comprehensive examinations by the parent population. The distribution of marks earned on the comprehensive examinations by the parent and sample populations are shown in Table IV.

TABLE IV
DISTRIBUTION OF MARKS RECEIVED ON THE COMPREHENSIVE
EXAMINATIONS BY THE PARENT AND SAMPLE POPULATIONS
JUNE, 1934

Marks Received	Parent Population of 600		Sample Population of 100	
	Number	Per Cent	Number	Per Cent
A	162	10	27	10
B	286	18	48	18
C	726	45	119	45
D	236	15	40	15
F	186	12	32	12
Total	1,596	100	266	100

The entries in the table show that the percentage distribution of marks is identical for both populations. The sample population used in the partial standardization of the diagnostic materials constructed by the investigator is thoroughly representative of the parent population, therefore, on the basis of the percentage of marks received on the comprehensive examinations, June, 1933.

In order to obtain a satisfactory statistical criterion of the goodness of fit or representative quality of the samples,

the Chi Square Test¹ was applied to all samples of 100 used in the partial standardization of the diagnostic materials. The results of the application of the Test to these samples is shown in Table V. The values of P obtained through the use of the Chi Square Test, as shown in the table, are very high owing to the typical or representative quality of the samples. The lowest value of P obtained on any test was .953719 and for any sub-test, .774566. Holzinger² states that "...unless the value of P be .2 or more, the fit cannot be regarded as good and other curves should be tried." The values of P obtained for the samples used in the study indicate a satisfactory degree of reliability for the sampling procedures.

The superior, middle, and inferior groups of 50.—One of the purposes of the study was to construct and partially standardize a diagnostic instrument in reading and study which would make possible the early differentiation between potentially superior and inferior students. To accomplish this purpose it was necessary to determine the differences which exist between superior and inferior students. Accordingly, three student samples were used in the further administration of the diagnostic materials to determine such differences. They were samples in only one sense of the word. They were total populations within their respective areas of performance but were samples of the total class of 700. These three samples consisted of 50 superior, 50 middle, and 50 inferior students selected on the basis of their performance on the comprehensive examinations. The term middle is used advisedly since these students were not average in the ordinary sense of the word. An average student might have received different combinations of A, B, C, D, and F marks on the comprehensive examinations. The middle students used in this study, however, were those who received only C's and at least two C's on these examinations. This group was used throughout the study to delimit the central area of the scale and to clarify the relative areas occupied by the superior and inferior students. The distributions of marks earned by these three groups on the comprehensive examinations are shown in Table VI.

¹Karl J. Holzinger, Statistical Methods for Students in Education, p. 245. Boston: Ginn and Company, 1928.

²Ibid., p. 247.

TABLE V

RESULTS OF THE CHI SQUARE TEST FOR GOODNESS OF FIT
OR REPRESENTATIVE QUALITY OF THE SAMPLES
USED IN PARTIALLY STANDARDIZING
THE EXPERIMENTAL MATERIALS

Type of Measure	(f)*	(f't)	(C)	(W)	(n)	(X ²)	(P)
Amer. Council							
1. Compl.	700	100	.14285	3.99	14	3.9	.991865
2. Art. L.	700	100	.14285	3.99	15	4.0	.995466
3. Anal.	700	100	.14285	1.99	11	3.5	.964385
4. Arith.	700	100	.14285	3.99	8	3.1	.874480
5. Oppos.	700	100	.14285	2.99	7	2.9	.819932
6. Total	700	100	.14285	9.99	22	5.6	.999523
Engl. Placement							
A. Spelling	650	100	.15384	0.99	6	2.2	.819314
B. Word.	650	100	.15384	0.99	9	3.0	.934357
C. Diction.	650	100	.15384	1.99	16	4.1	.997176
D. Sent. G.	650	100	.15384	0.99	9	3.2	.918910
E. Sent. F.	650	100	.15384	1.99	12	3.6	.978241
F. Parag.	650	100	.15384	0.99	7	2.7	.842103
Eng. Total	650	100	.15384	1.99	12	3.6	.978241
Iowa Reading							
1A. Science	664	100	.15060	1.99	7	2.6	.853188
1B. Literature . . .	664	100	.15060	1.99	7	2.4	.875358
2A. Soc. Science . .	664	100	.15060	0.99	7	2.8	.831017
2B. Science	664	100	.15060	0.99	6	2.3	.804398
2C. Math.	664	100	.15060	0.99	6	2.4	.789482
2D. English	664	100	.15060	0.99	7	2.7	.842103
3A. Cent. Id. . . .	664	100	.15060	0.99	6	2.5	.774566
3B. Outl.	664	100	.15060	0.99	6	2.4	.789482
4. Sent. M.	664	100	.15060	1.99	10	3.3	.948430
5A. Use Ind. . . .	664	100	.15060	0.99	7	2.6	.853188
5B. Key Wo.	664	100	.15060	0.99	8	3.2	.863957
Total Comp. . . .	666	100	.15015	4.99	24	5.8	.999036
Total Rate	666	100	.15015	1.99	22	5.4	.999878
Iowa Total	664	100	.15060	9.99	16	4.2	.996615
Pers. Schedule . .	679	100	.14727	4.99	21	5.3	.999477
Minn. Read. V. . .	647	100	.15455	4.99	16	4.0	.999830
Minn. Read. P. . .	647	100	.15455	1.99	10	3.2	.953719
Minn. Total	647	100	.15455	4.99	12	3.5	.980321
H. S. Average . . .	600	100	.16666	1.99	12	3.4	.982402
H. S. Rank	600	100	.16666	4.99	14	3.7	.993214
Marks on Comp.							
Examination . . .	600	100	.16666		17	4.0	.999830

*Legend. (f) Total number in parent population
(f't) Total number in sample population
(C) Constant used to enter each frequency interval
(W) Width of class interval
(n) Number of class intervals
(X²) Chi Square function
(P) Value of X²

TABLE VI
DISTRIBUTION OF MARKS RECEIVED ON THE COMPREHENSIVE
EXAMINATIONS, JUNE, 1933, BY THE SUPERIOR,
MIDDLE AND INFERIOR GROUPS

Marks	Superior Group		Middle Group		Inferior Group	
	No.	Per Cent	No.	Per Cent	No.	Per Cent
A	121	79				
B	32	21				
C			140	100		
D					12	10
F					108	90
Total .	153	100	140	100	120	100

The 50 superior students were those who received either three A's or two A's and one B, and the 50 inferior students were those who received at least two F's. The data presented in Table VI show that the three samples were mutually exclusive in terms of their performance on these examinations. This quality of being mutually exclusive increased the differential value of the responses obtained on the diagnostic materials administered to the three groups.

The foregoing discussion of sampling methods shows specifically the nature and statistical reliability of the samples used in the study. These considerations will not be discussed in detail in later chapters.

. General Statistical Methods

Certain common statistical measures were used throughout the study. The major type of statistical treatment of the data consisted of the computation of a measure of central tendency, the ordinary statistical average or arithmetical mean. This was used because of its simplicity, statistical reliability,¹ and understandability for administrative officers. The computation of this measure and of the more infrequently used measures such as percentages and probable errors of the means were made in con-

¹Ibid., p. 84.

formity with the standard methods indicated in treatises in statistics, especially that of Holzinger.

The average marks in the core-subjects taken in high-school by the students used in the study were brought to a common basis by projecting all marks on a scale ranging from 75 to 100. Such a procedure is recommended and used by the Admissions office of the University of Chicago. This method of equalizing marks served to smooth the differences existing between schools having different passing marks or grading systems. Similarly, letter marks were projected upon the scale under the assumption that such marks represent a normal frequency distribution.

The common method of computing percentile distributions, as indicated by Holzinger, was followed in all cases excepting the distribution of high-school ranks. In computing rank in a high-school graduating class, the method which is commonly used by college administrators and personnel officers was followed, both because it is the common procedure and because it was used by Reitz¹ in the construction of his prediction formulas which are applied, in part, in the experimental portion of this study. In the computation of these ranks the following formula was used:

$$\text{Rank} = (100 - \frac{\text{rank of student}}{\text{number of class}})$$

Hull² has recommended that the following formula be used in future computations of high-school rank:

$$\text{Per cent position} = \frac{100 (R - .5)}{N}$$

where R is the rank of the particular student and N is the total number of students in the graduating class. The per cent position is then transmuted by means of Hull's tables into a linear score on a 100-point scale. Distributions computed by the latter method approach normality to a considerable extent while distributions computed by the former method tend to result in a piling-up of the ranks at the upper end of the distribution.

¹William Reitz, "Forecasting Success of Students under the New Plan at the University of Chicago," Unpublished report to the Dean of Students and University Examiner, University of Chicago, May, 1933.

²Clark L. Hull, Aptitude Testing, p. 388. Chicago: World Book Company, 1928.

The statistical measures and methods that have been described are the principal ones used in the treatment of the data secured in the study. They will not be discussed again when the construction and partial standardization of the diagnostic materials are described in later chapters.

CHAPTER III

THE CONSTRUCTION AND PRELIMINARY USE OF THE DIAGNOSTIC MATERIALS

Problem of the Chapter

The primary purpose of the study, as reported earlier, is the construction and partial standardization of a comprehensive diagnostic instrument in reading and study at the college level, which makes possible the early differentiation between potentially superior and inferior students, for purposes of subsequent remediation.

The development of the diagnostic instrument involved, however, a preliminary analysis of all previous investigations of reading and study abilities to determine the necessary elements of a complete diagnosis. As a result of this analysis, the essential diagnostic materials were either adapted from other sources or constructed by the investigator. The actual use of the diagnostic instrument involved the development of comparative norms by means of which the differentiation or classification of superior and inferior students might be accomplished. The problem of this chapter, therefore, is to describe the construction and partial standardization of the diagnostic materials which compose the instrument.

Description and Administration of the Diagnostic Materials

Several hundred books, periodical articles, and scientific investigations on various aspects¹ of reading and study abilities² have been published over a period of years. This mass of literature furnished a wealth of diagnostic and remedial suggestions, case histories, and facts which showed the relationship between reading and study abilities and general scholastic success. A major problem in the development of the diagnostic technique was

¹Gray, op. cit.

²Cowley, op. cit.

the careful survey and critical analysis of this body of literature to determine the significant elements or aspects of reading and study which should be measured by a comprehensive diagnosis. As a result of this analysis, the various phases of reading and study which previous studies had shown to be of considerable diagnostic and remedial significance were selected as the focal points around which diagnostic measures might be constructed or adapted.

It is apparent, therefore, that the contribution of the study lies not so much in the actual determination of the relationships which exist between the various aspects of reading and study abilities and scholastic success. The contribution rests rather in the identification of these relationships in previous studies and the integration of the diagnostic and remedial suggestions of other investigators into a comprehensive diagnostic instrument. The suggestions of previous investigators were, however, subjected to further empirical investigation to determine their actual value.

Considerable evidence was found during the analysis which pointed to the fact that an intimate relationship exists between scholastic success,¹ reading and study abilities,² and certain factors such as intelligence, personality, environment, physical and mental hygiene, memory, vision, association, hearing, and many others. Since the major purpose of the study was to develop an instrument by means of which a comprehensive diagnosis might be obtained, it seemed desirable to include in the diagnostic battery exploratory measures of these related factors.

As the analysis progressed, it became increasingly evident that the existing diagnostic measures at the college level were of the survey type which gives only a general diagnosis. This type is defined by Brueckner and Melby as including "....any pro-

¹Raymond D. Meade, "An Evaluation of the First Year under the New College Plan of the University of Chicago, Based upon Student Progress and Student Opinion." Unpublished Master's Thesis, Department of Education, University of Chicago, 1933.

²H. H. Remmers, "A Diagnostic and Remedial Study of Potentially and Actually Failing Students at Purdue University." Bulletin of Purdue University, Vol. XXVIII, No. 12. Studies in Higher Education, IX. Lafayette, Indiana: Purdue University Press, 1928.

cedure that may be used to secure general information regarding the general status of a pupil or of a class." The same writers state further:

"The chief limitation of this kind of diagnosis is that it does not give the teacher any facts on the basis of which to determine the specific causes of superiority or inferiority....These tests serve admirably as a means of securing a general measure of the pupil's present status....but they do not supply the kind of detailed information needed to make a diagnosis of a particular pupil's deficiency; nor can the teacher determine on the basis of the test results the particular phases of the subject in which deficiencies exist. However, these tests do help the teacher to select those pupils whose performances are not up to standard, and whose work should be more carefully studied to determine what steps should be taken to remedy the condition."¹

In view of the limitation of the use and value of general diagnostic measures, it became necessary for the investigator to construct various diagnostic devices which would reveal specific items of deficiency. To meet this need a further analysis of the literature was made to obtain valuable suggestions which might be incorporated into a series of reading and study check lists. As a result of this analysis, 931 of the more specific items recommended by specialists as good procedures were selected and organized into 19 reading and study check lists. Using this series of lists as the core of the diagnostic instrument, various other tests and measures were adapted from additional sources or were constructed by the investigator to measure those abilities or factors which the analysis of literature had shown to be significantly related to reading and study.

In order to provide a preliminary over-view of the diagnostic materials which were partially standardized in the study, a complete list is presented at this point. The reasons for the selection and use of each of these diverse measures are discussed later.

¹Brueckner and Melby, op. cit., p. 124.

Materials Which Compose the Comprehensive
Diagnostic Instrument

- A. The Iowa Silent Reading Test
 - 1. Paragraph Meaning
 - a. science
 - b. literature
 - 2. Word Meaning
 - a. social science
 - b. science
 - c. mathematics
 - d. English
 - 3. Paragraph Organization
 - a. selection of central idea
 - b. outlining
 - 4. Sentence Meaning
 - 5. Location of Information
 - a. use of the index
 - b. selection of key words
 - 6. Total Comprehension Score
 - 7. Rate of Silent Reading
 - 8. Total Score on Iowa Reading Test
- B. Minnesota Reading Examination
 - 1. Vocabulary
 - 2. Paragraph Reading
 - 3. Total Score
- C. The University of Chicago Placement and Qualifying Examination in English Composition. September 1932
 - 1. General Usage
 - a. spelling
 - b. wordiness
 - c. diction
 - d. sentence groups
 - e. sentence faults
 - f. paragraphing
 - g. total score
- D. The American Council on Education Psychological Examination, 1932
 - 1. Completion
 - 2. Artificial Language
 - 3. Analogies
 - 4. Arithmetic
 - 5. Opposites
 - 6. Total Score
- E. The Thurstone Personality Schedule, 1929 Edition
- F. High-School Factors
 - 1. Average of High-School Marks
 - 2. Rank in High-School Graduating Class
- G. The Comprehensive Reading and Study Check Lists
 - 1. General Physical Factors of Reading, Study, and Hygiene
 - 2. Making Notes. (General, Reading, Lecture, Laboratory)

3. Outlining and Summarizing
4. Making Bibliographies and Preparing Written Assignments
5. Using the Libraries
6. Attending Lecture and Discussion Sections
7. Working in the Laboratories
8. General Study Factors
9. Building a Vocabulary
10. Preparing Reading Assignments and Using Text-Books and Reference Books
11. Special Factors of Reading Rate
12. Special Factors of Reading Comprehension
13. General Reading Factors
14. Memorizing
15. Reviewing and Preparing for Examinations
16. Taking Examinations
17. Developing and Maintaining Interest in Your Work
18. Thinking and Problem-Solving
19. Factors of Habit Formation

H. The Battery of Informal Tests

1. Reading Rate (number of seconds required to complete the selection)
 - a. normal (silent)
 - b. rapid (silent)
 - c. normal (oral)
2. Observation of Eye-Movements while Reading
 - a. average number of fixations per line
 - b. average duration of fixation (in 25th of a second)
 - c. average number of regressive movements per line
 - d. accuracy of return sweeps
3. Pronunciation
4. Vision (distance in feet from newspaper clipping) (Caslon type)
 - a. 72-point boldface
 - b. 24-point medium
 - c. 14-point condensed
 - d. 14-point boldface
 - e. 12-point boldface
 - f. 7-point
5. Eye-Voice Span (number of letter-spaces)
6. Recognition Unit (short exposure-number of letter spaces-without eye-movement)
7. Recognition Unit (short exposure, number of letter-spaces, with eye-movement)
8. Recognition Unit (number of letter-spaces while reading)
9. Proof-Reading (finding typographical errors)
10. Speed of Association (controlled word)
11. Speed, Accuracy, and Coordination of Motor Reaction, and Attention
 - a. tapping
 - b. card-sorting
12. Attention
 - a. cancellation of letters
 - b. dot-counting

- c. duration of visual attention
- 13. Digit Memory-Span (auditory forward, auditory reverse, and visual forward)
- 14. Rote Memory on Nonsense Syllables
- 15. Visual Memory and Grasping of Essential Facts
- 16. Auditory Factual Memory and Grasping of Essential Facts
- 17. Aussage Observation and Memory (objects in picture)
- 18. Observation and Memory (examiner's person and attire)
- 19. Hearing
- 20. Cause and Effect Relationships
- 21. Following Test Directions

I. Scholastic Factors

J. Personal Traits and Characteristics

K. Personal Appearance, Health, and Hygiene

The diversity of the diagnostic materials which compose the partially standardized portion of the instrument is readily apparent from a consideration of the foregoing list. Using this list as a point of reference, each type of diagnostic material is described in succeeding sections of the chapter. The comparative norms derived from the administration of this portion of the diagnostic instrument will not be presented with the description of each type of material. They will be reported in detail in a later section of the chapter which describes the construction of the comprehensive and classificatory rating scales.

The placement tests and admissions data.--The Board of Examinations of the University of Chicago administers a battery of placement tests to each entering freshman class. This battery includes tests in reading, English, intelligence or scholastic aptitude, and personality. The scores earned on these tests are used, subsequently, as one of the major bases for the guidance program of the college advisory staff.

These tests were selected by the Board of Examinations after a careful consideration of all of the available tests in each field. They meet the generally accepted criteria of a good test, such as ease of administration and scoring, objectivity, validity, reliability, tentative norms, and utility.¹

¹G. M. Ruch, The Objective or New-Type Examination, pp. 27-66. Chicago: Scott, Foresman and Company, 1929.

The consideration which led the investigator to adapt these tests was the fact that they give a general survey of the relative status of students in each ability or test, and, in addition, the various sub-tests and items of which they are composed are of considerable diagnostic value. For example, the Iowa Reading Test is composed of 11 sub-tests, each of which provides a measure of some specific aspect of reading and study ability such as rate of reading, reading comprehension, use of the index, selection of the central idea, outlining, selection of key words, vocabulary, and paragraph meaning. The other placement tests contain sub-tests or items of similar diagnostic significance in their respective fields which contribute information of considerable value to the comprehensive diagnosis which the instrument attempts to provide. It has been shown further that these tests bear a positive relationship to scholastic success as well as to reading and study abilities.¹

A further consideration which influenced the investigator to include these tests in the diagnostic instrument was the fact that they are administered each autumn by the Board of Examinations and the scores are made available shortly thereafter. The time and labor involved in the administration of the complete diagnostic instrument was lessened considerably by this fact.

The exact titles of these placement tests follow: (1) the Iowa Silent Reading Test; (2) the Minnesota Reading Examination for College Students; (3) the University of Chicago English Placement Test and Qualifying Examination, 1932 edition; (4) the American Council on Education Psychological Examination; and (5) the Thurstone Personality Schedule. The construction and administration of these tests are described fully in the Manuals which accompany them.²

¹Meade, op. cit.; Reeves and others, op. cit., IV-V.

²H. A. Greene, A. N. Jorgensen, and V. H. Kelley, Manual of Directions. (To accompany the Iowa Silent Reading Test.) Chicago: World Book Company, 1931. Melvin E. Haggerty and Alvin C. Eurich, Examiner's Manual. (To accompany the Minnesota Reading Examination for College Students.) Minneapolis, Minnesota: The University of Minnesota Press, 1930. M. W. Richardson, J. T. Russell, J. M. Stalnaker, and L. L. Thurstone, Manual of Examination Methods. The University of Chicago Board of Examina-

The adaptation of these tests to the specific purposes of the present study, that is to the diagnosis of and differentiation between potentially superior and inferior students, necessitated the derivation of local norms or standards of comparison since the student body of the University of Chicago is of a distinctly higher quality than that in many other institutions. This is shown by the fact that the mean difference between the local and national median scores on the Psychological Examination for the past six years, 1928-33, is 51.3 score points in favor of the University of Chicago students. The 50 inferior students used later in the study had a mean local percentile rank on this test of 16.8 and a median raw score of 164. The national median and the 50 percentile for the same test, 1932 edition, were also 164.¹ The lower group of students at the University of Chicago is equivalent in intelligence, therefore, to the average group at many other institutions, while the Chicago student body as a whole is distinctly superior to that of many other institutions.

To meet this need for local comparative norms the frequency distributions of the raw scores and percentile ranks of the freshman class of 1932-33 were obtained from the Board of Examinations for each of these tests. The scores used by the investigator in determining the local differential and diagnostic norms for the comprehensive and classificatory rating scales were secured from these distributions by the statistical procedures described in detail in Chapter II.

Certain differences were found to exist between the performance of superior and inferior students on this battery of

tions, Preliminary edition, June 1933. (For additional information concerning the English test, September 1932, write to the English Department, University of Chicago, Chicago, Ill.)

L. L. and Thelma Gwinn Thurstone, Manual of Instructions. (To accompany the Psychological Examination for High-School Graduates and College Freshmen.) Washington, D.C.: The American Council on Education, 1932. L. L. and Thelma Gwinn Thurstone, Manual to Accompany the Personality Schedule. (Printed folder of instructions accompanying each package of tests.) Chicago: University of Chicago Press, 1930.

¹Thurstone and Thurstone, Manual of Instructions, op. cit.,

TABLE VII
GENERAL STATUS OF THE SUPERIOR AND INFERIOR STUDENTS ON THE
UNIVERSITY OF CHICAGO PLACEMENT TESTS, SEPTEMBER 1932

Measures	Psychological Examination		Iowa Reading Rate		Iowa Reading Comprehension		Personality Schedule		English Placement	
	Sup.	Inf.	Sup.	Inf.	Sup.	Inf.	Sup.	Inf.	Sup.	Inf.
Mean Percentile Rank	77.6	16.8	62.0	31.4	75.1	20.8	55.2	43.1	72.9	22.3
Mean Raw Score . .	261.3	164.1	36.2	27.6	179.7	135.5	37.1	29.4	29.0	22.0
Range Percentile Rank	20-100	0- 95	2-100	0-100	7-100	0- 80	0- 99	0- 96	7-100	0- 88
Range Raw Score .	169-365	86-309	17- 61	12- 61	115-210	69-182	2-102	5- 79	19- 40	19- 32

tests which are of general interest. A preliminary presentation of these differences is made, therefore, in Table VII.

A comparison of the entries in this table shows that there is considerable difference in the performance of the two groups on each of the placement tests. The larger scores and higher ranks are all earned by the superior group. The mean differences in percentile ranks on the five measures, as computed from the data in the table, are as follows: intelligence, or scholastic aptitude, 60.8; reading rate, 30.6; reading comprehension, 54.3; personality, 12.1; and English, 50.6.

The larger score earned by the superior group on one of these measures may be considered as unfavorable, however, since it indicates a greater degree of mal-adjustment. That measure is the Personality Schedule, which is based upon the degree of adjustment or mal-adjustment present. This tendency of the superior students to be relatively less well adjusted has been noted in a previous study.¹

The admissions data on high-school average and rank in high-school graduating class were included in the diagnostic instrument because previous studies² have indicated, without exception, that they are of considerable significance in predicting scholastic success and in differentiating between superior and inferior students. These data were computed by the statistical procedures described in Chapter II from the information submitted by the high-school principals in the application for admission blanks. The resulting norms were included under section F of the diagnostic instrument to increase the differential value of the diagnosis and to furnish supplementary information concerning the students' high-school background.

Certain differences were found between the status of superior and inferior students on these two measures which are also of general interest. The superior group had a high-school average of 93.2 and a high-school rank of 95.7, as against a high-school average of 85.1 and a high-school rank of 65.1 for the inferior group. The range in high-school average was 85 to 96 for the superior groups and 70 to 95 for the inferior group, while the range in high-school rank was 75 to 99 for the superior group and

¹Meade, op. cit.

²Reeves and others, op. cit., V., 70.

14 to 98 for the inferior group. It is apparent, therefore, that the superior group made higher averages and ranks and were more homogeneous as indicated by the lesser degree of variability in the range on the two measures.

For the future use of this portion of the diagnostic instrument the following procedure was adopted: (1) the actual administration of the placement tests should be left with the Board of Examinations; (2) the raw scores and percentile ranks of entering freshmen on these tests should be made available by the Board as soon as possible after their administration, in addition to the high-school averages and high-school ranks; and (3) the actual test papers should be available for the further and more specific diagnosis of inferior students by the properly authorized remedial workers.

The tests and measures described thus far comprise the more formal or standardized portion of the diagnostic instrument and have been used in many similar studies. The diagnostic materials which comprise the major contribution of the study and which differentiate the present technique from those developed in previous studies are presented in the succeeding sections of the chapter.

The comprehensive reading and study check lists.--The primary purpose of a diagnosis is to determine or to identify the nature of a student's deficiency. As reported earlier, this purpose is achieved only in a very general manner by the use of the standardized placement tests administered by the Board of Examinations. Before the proper remedial procedures can be determined the remedial worker must have a detailed knowledge of the specific items of deficiency in addition to a knowledge of the general status of the reading and study abilities of particular students. This type of diagnosis may be termed the "analytical" diagnosis, which, according to Brueckner and Melby,¹ is "...based upon a detailed analysis of the several skills which constitute the general ability under consideration."

An analysis of previous investigations and published tests revealed the fact that no measure or instrument was available, at the college level, which would provide such a diagnosis. To meet this need a critical analysis was made of all available

¹Op. cit., p. 125.

literature on reading and study, diagnosis and remediation, and factors of hygiene, to determine the specific items which would give a complete diagnosis of reading and study abilities, habits, or activities. As a result of this analysis, 931 of the more specific suggestions were summarized and used in the construction of 19 reading and study check lists. The list of 19 titles, as reported on pages 25 and 26, indicates that the items cover a wide range of reading and study activities and of the factors of hygiene which are of value in the academic life of college students.

Taken as a whole, the check lists provide a rather complete guide to the best reading and study activities, as established through previous studies, which students should know and use in order to obtain maximal efficiency in these two fundamental types of academic activity. The items which compose the lists are all good procedures, as recommended by specialists, and are stated in positive form. In order to give a more adequate picture of the type of items which are included in this portion of the diagnostic instrument, a section of Check List number 16, "Taking Examinations," is reproduced at this point.

1. Get your normal amount of sleep the night before the examination.
2. Take a light breakfast or lunch of easily digested food before the examination.
3. Be on time for the examination.
4. Have all necessary materials with you.
5. Be physically and mentally alert; concentrate actively upon the examination.
6. Trust your memory; maintain an attitude of confidence toward the results of the examination.
7. Take the test seriously; avoid being over-confident.
8. Follow directions exactly.
9. Try for as high a score as possible; do yourself justice on the test.
10. Read the instructions and questions carefully.
11. Re-read the answer to each question before passing on to the next question.
12. If time is available, look back over the test and revise if necessary.
13. Be explicit; do not expect the examiner to take anything for granted.
14. Be neat and careful in all of your written work.

There are 38 suggestions of a similar nature in this Check List which, as a group, give a fairly complete analysis of the activities involved in the particular type of study to which it refers.

In order to obtain a more specific knowledge of the actual degree of use of an activity item than is possible with the simple

yes-no type of response, the following five-point scale of responses was adopted:

1	2	3	4	5
never	seldom	often	usually	always

Total deficiency on any item is indicated by the negative response of never, while lesser degrees of deficiency are shown by the responses seldom and often. The responses usually and always are interpreted as being of high quality, indicating no deficiency on the particular item. Since the items are all good procedures and are stated in positive form, the diagnostic value of the five-point scale is apparent.

The 19 reading and study check lists, as thus developed, may be used to achieve two purposes: (1) to obtain a comprehensive and specific diagnosis of deficiencies in the chief aspects of reading and study activity; and (2) to provide an elaborate set of remedial guides to the good procedures that should be followed in obtaining maximal efficiency in reading and study.

The method used in administering the check lists was oral, personal, and individual. The investigator asked each subject leading questions concerning his use of the reading and study activities referred to in the lists and checked the proper responses on a master data sheet. The lists served to control the interview and to reduce all responses to comparable terms. Some of the major defects¹ of the ordinary interview and questionnaire methods were overcome by this method of administration. This precaution was imperative in establishing the most adequate norms for the partial standardization of this portion of the diagnostic instrument. The fact should be pointed out, however, that the individual method of administration may be changed quite readily to the impersonal group administrative method by the simple procedure of placing the words "Do you" before each item or each group of items in the check lists and presenting them to the students in printed form.

The derivation and statistical significance of the comparative norms developed for these lists are discussed in detail in a later section of the chapter which describes the construction of the comprehensive and classificatory rating scales. Certain

¹Leonard V. Koos, The Questionnaire in Education, pp. 144-67. New York: The Macmillan Company, 1928.

general differences which were found between the performance of superior and inferior students on the lists are of some interest. These differences are presented in preliminary form in Table VIII.

TABLE VIII
RESULTS OBTAINED FROM THE PRELIMINARY ADMINISTRATION
OF THE READING AND STUDY CHECK LISTS TO THE
SUPERIOR AND INFERIOR GROUPS

Check List Letter and Rating Scale Number	Total Score of the Aver- age Student		Per Cent of Total Pos- sible Score		Per Cent of Superiority for the Su- perior Group	Difference in Per Cent of Total Possible Score in Fa- vor of the Su- perior Group
	Sup.	Inf.	Sup.	Inf.		
A 1 . .	392	374	74.0	70.6	4.8	3.4
B 2 . .	267	227	71.2	60.5	17.6	10.7
C 3 . .	130	110	66.6	56.4	18.2	10.2
D 4 . .	202	160	74.8	59.3	26.3	15.5
E 5 . .	93	79	84.5	71.8	17.7	12.7
F 6 . .	108	87	77.1	62.1	24.1	15.0
G 7 . .	131	109	77.1	64.1	20.2	13.0
H 8 . .	294	228	74.4	57.7	28.9	16.7
I 9 . .	81	69	60.0	51.1	17.4	8.0
J 10 . .	202	162	74.8	60.0	24.7	14.8
K 11 . .	133	107	68.1	50.9	24.3	17.2
L 12 . .	174	140	66.9	53.8	24.3	13.1
M 13 . .	226	188	76.6	63.7	20.2	12.9
N 14 . .	182	164	79.1	71.3	11.0	7.8
O 15 . .	247	219	86.7	76.8	12.8	9.9
P 16 . .	173	156	91.1	82.2	10.9	8.9
Q 17 . .	133	111	80.6	67.3	19.8	13.3
R 18 . .	222	179	79.3	63.9	24.0	15.4
S 19 . .	129	115	75.9	67.6	12.2	8.3
Total	3519	2984			359.4	227.7
Mean			75.6	63.7	18.9	11.9
Range			60.0- 91.1	50.9- 82.2	4.8- 28.9	3.4- 17.2

The entries in this table show that the differential value of the various lists is subject to considerable variation. The mean difference in per cent of the total possible score on all lists is 11.9, in favor of the superior group. The range of differences in these same percentages in favor of the superior group, however, is from 3.4 to 17.2. The general significance of these differences is shown to better advantage by the percentage of superiority which the superior group possesses over the inferior group. The mean per cent of superiority of this group is 18.9, while the range in superiority is from 4.8 to 28.9. As shown by the per cent of total possible score achieved by these two groups, the superior students operate at only 75.6 per cent efficiency and the inferior students at only 63.7 per cent efficiency. It is apparent that both of these groups could increase the efficiency of their reading and study activities to a considerable degree.

The reading and study check lists described in this section of the chapter compose one of the major contributions of the study. They form the core around which the supplementary materials are developed and with which they are integrated. The nature of these supplementary diagnostic materials are considered next.

The battery of informal individual tests.--There are certain aspects of reading and study abilities which do not lend themselves to ordinary types of group measurement. Adequate vision, for example, is a fundamental sensory pre-requisite for efficient reading and study. Similarly, hearing ability is necessary for the effective performance of certain study activities, such as attending lectures or class discussions. Memory is another ability which is fundamental in ideational types of learning, including reading and study. An adequate diagnostic instrument should include, therefore, some measure of these related abilities as a means of determining possible causes of reading and study deficiencies.

To meet this evident need, the literature on psychological and physical measurements was reviewed for the purpose of obtaining suggestions on the measurement of these abilities. As a result of this analysis, 21 tests were constructed by the investigator which give a direct yet simple measure of these abilities. For example, the test of vision is administered by tacking an ordinary newspaper clipping, which is printed in various sizes

of type, on a wall and instructing the subject to stand at the point farthest away from the clipping at which he can still read and interpret the exposed line of print. The score for the test is the distance in feet between the wall and the subject's eyes. A similarly direct measurement is made of hearing ability. The examiner, using an ordinary wrist-watch, takes four measurements of the distance at which the subject can no longer hear the ticking of the watch as the examiner moves away from him. The measurements are as follows: one from the left side, one from the right side, one from in front, and one from in back of the subject.

The tests of memory are of a similarly direct nature, although several of these measurements are needed to test the various aspects of memory ability. For example, the subject reads lines of digits silently and recalls them serially, to test visual memory for digits. The subject reads other lines of digits orally and recalls them serially, to test verbal memory for digits. Pure memory of rote materials, which have no logical connection with each other, is tested by having the subject read a list of non-sense syllables twice and recall them serially. Visual memory of essential facts is measured by having the subject read silently a paragraph which describes the requirements of a magazine contest and recall these requirements after the reading is completed. A similar measure of his aural memory for facts is obtained through having the examiner, himself, read orally 13 statements concerning the facilities which a certain hotel offers, and asking the subject for verbatim recall. A further type of visual memory is tested by having the subject look at a picture of an ordinary living room of a home for a minute and a half and recall all of the objects that he has observed in the picture, after it is removed. A subject's general observation and memory is tested by having him close his eyes and answer certain questions concerning the examiner's attire. Through these tests, an attempt is made to obtain a diagnostic knowledge of the various aspects of a subject's memory ability.

Similar direct measurements are obtained of various aspects of reading and study abilities, such as normal oral reading, pronunciation, normal silent reading, rapid silent reading, attention, and speed of word association, which various investiga-

tors¹ have shown to be of value in diagnostic and remedial work. The complete list of test titles, as reported earlier, indicates the nature and extent of this portion of the diagnostic instrument.

The battery is designed for use as an individual rather than a group measure. It is to be used as a supplementary form of diagnosis after the general administration of the placement tests and of the reading and study check lists.

The tests were partially standardized by being administered, individually, to the typical sample of 100, the 50 superior, 50 middle, and 50 inferior students, described in Chapter II. The norms derived from this administration are presented in a later section of the chapter in which the development of normative methods for the diagnostic materials is discussed.

When used in conjunction with the placement tests and check lists, the battery of individual tests completes that portion of the diagnostic instrument which is aimed primarily at the determination of the present status of a student in reading and study abilities. One additional element of the partially standardized portion of the diagnostic instrument has not been described, namely, the informal personal ratings.

The informal personal ratings.--Previous studies² have shown that the student's general industry, punctuality, personality traits, determination, cooperation, physical condition and hygiene, economic and family background, and many other factors are of considerable importance, either causal or concomitant, in determining or conditioning scholastic success as well as reading and study achievement.

No diagnostic instrument which is designed for use in individual case studies would be complete or adequate, therefore, without obtaining some estimate or measure of the personal traits and general background of the individual student. An analysis was made of the literature on reading and study of the case-history

¹Arthur Edwin Traxler, "The Measurement and Improvement of Silent Reading at the Junior-High-School Level." Unpublished Doctor's Dissertation, Department of Education, University of Chicago, 1932. Also Gray, op. cit.

²Remmers, op. cit.; Gray, op. cit. Marion Monroe, Children Who Cannot Read. Chicago: The University of Chicago Press, 1932.

type to determine the traits or factors which were of significance in conditioning reading and study achievement and scholastic success. It was found that many of these factors are relatively intangible in nature and do not lend themselves to measurement by ordinary testing procedures. A rating scale device was constructed by the investigator, therefore, to meet this need. As a result of the analysis of the literature and the identification of the factors which previous studies had shown to be of most significance, those items or traits were selected which would lend themselves more adequately to rating purposes. These were grouped under three headings: personal traits or characteristics; scholastic factors; and general personal appearance, health, and hygiene. Under the heading personal traits were included such items as enthusiasm, punctuality, cooperation, cheerfulness, tact and social poise. Under scholastic factors were included such items as attitude toward teachers, reaction toward criticism, general attention, use of time, quality of notes, and quality of written assignments. The heading personal appearance, health, and hygiene includes such items as absence of unpleasant body odor, general physical neatness, physique, posture, and appearance of skin.

Since these ratings are essentially subjective, an attempt was made to select items or traits which were relatively more objective in nature, or upon which the ratings of other people were available. The investigator did not presume to make these ratings on the basis of his knowledge and opinion alone. The final ratings given to students on these traits were a composite of ratings derived from many sources. These sources were the statements of the student in the application for admission blank, ratings by high-school teachers and principals, personnel reports in the office of the Dean of Students, instructors' reports, advisers' reports, medical records, analysis of the responses made by the student to the individual items of the Personality Schedule, and observations made by the investigator during the administration of the diagnostic materials.

After a careful and critical consideration of the ratings, personal estimates, and observations obtained from these sources, a composite rating was given by the investigator. A common five-point form of rating was used as the basis for these rating scales:

1	2	3	4	5
very poor	poor	fair	good	very good

The ratings were obtained for the same groups of students used in the administration of the tests and check lists. The norms derived are presented in the succeeding section of the chapter which deals with the development of normative methods.

The ratings are, in the final analysis, a subjective type of measure. No adequate diagnosis or understanding of an individual student could be achieved, however, without obtaining some knowledge or estimate of these traits. This element of the diagnostic instrument was retained and used in subsequent diagnosis, therefore, in preference to leaving these areas of a student's life and personality totally unexplored.

Development of Comparative Norms for the Diagnostic Materials

Owing to the extreme diversity of the diagnostic materials, as shown in the foregoing description, it was necessary to use a statistical device for reducing the scores earned by students on the various measures to some simple, comparable, and understandable basis for purposes of subsequent summation and combination. The use of a complex system of standard scores which required considerable computation would defeat the very purpose of the diagnostic instrument since little use would be made of such an instrument by heavily burdened counselors and administrative officers. To meet this need a Comprehensive Rating Scale was constructed by means of which the scores of a student on all of the diverse types of diagnostic materials could be reduced to a five-point basis.

Since one of the major functions of the diagnostic instrument is to obtain a preliminary classification of or differentiation between potentially superior and inferior students, it was necessary to develop further comparative norms for this purpose. To meet this need a three-point Classificatory Rating Scale was constructed, composed of the mean rating scale scores on all diagnostic materials of the 50 superior, 50 middle, and 50 inferior students. Through the use of this device entering students could be classified into the proper categories, for purposes of subsequent remediation.

The construction of these two scales is described therefore in the succeeding sections of the chapter.

The comprehensive rating scale.--Raw scores on test materials have little meaning except in relation to norms of comparison. For example, a raw score of 45 on a test would indicate a high degree of ability when the total possible score on that test is 50 and a low degree of ability if the total possible score on that test were 300. To meet the need for reducing the scores on all of the diagnostic materials to simple, comparable, and understandable terms the investigator made a preliminary investigation of various systems of standard scores. The common standard scores, as reported by Holzinger,¹ are computed by subtracting the mean score of the group from the raw score of the individual and dividing this remainder by the standard deviation. The resulting standard score shows the relative position of a student in the group in terms of a number of standard deviations above or below the mean. Owing to the considerable amount of computation which is involved in the derivation of each of these scores and the difficulty of their interpretation, the investigator conferred with various personnel officers in several colleges to determine whether they would make use of an instrument whose scores were reduced to these terms. The unanimous opinion was to the effect that such scores would impose a considerable limitation upon the general administrative use of such materials.

In order to satisfy the demand for a simple statistical device which could be used without elaborate statistical treatment or interpretation, a rating scale device was constructed which reduced all scores on the diagnostic materials to a comparable five-point basis. The rating scale score shows the relative position of an individual raw score in terms of the total distribution of the class of 700 students, 1932-33, on that test.

Since the complete diagnostic instrument was too lengthy to be administered to the total class of 700 students, carefully selected typical samples were obtained from this total population by the statistical procedures described in detail in Chapter II. The high values of P, .999830 to .774566, obtained by the application of the Chi Square Test to these samples indicated that the samples were statistically reliable.

¹Op. cit., p. 119.

Through the administration of the diagnostic materials to these student samples, 100 scores were obtained on each test, sub-test, check list, and group of personal ratings, which gave a typically representative sample of the parent population on these measures. These individual scores were used as the basis upon which the Comprehensive Rating Scale was constructed. The actual construction of the scale was accomplished by ranking the 100 scores on each type of diagnostic material in ascending order of ability and isolating the 1st, 20th, 40th, 60th, 80th, and 100th individual scores. These six values formed a continuous five-point scale, as shown in Table IX.

The scores which compose this scale are the basic norms for the diagnostic instrument. Through the use of this rating device, three purposes are served: (1) all scores on the different diagnostic materials can be reduced to a comparable five-point basis; (2) a general estimate can be obtained of the status of a student on these diagnostic materials, since the five-point rating has a spread which covers the entire range of scores in the parent distribution of 700 students. For example, a rating of 1 could be classified roughly as very low or very poor, a rating of 2 as poor, a rating of 3 as fair or average, a rating of 4 as good, and a rating of 5 as very good; and (3) a final rating scale score can be derived for the entire diagnostic instrument or for any portions of it by summing or combining scores on this comparable five-point basis.

The third of these purposes is of especial significance, at present, since that use of the comprehensive rating scale was involved in the construction of a second form of scale by means of which students could be definitely classified as potentially superior or inferior on the diagnostic materials.

The classificatory rating scale.--Through the use of the comprehensive scale, a knowledge of the general status of students on the diagnostic materials may be obtained readily, as shown in the foregoing discussion. The norms on this scale can not be used, however, to determine whether the degree of deficiency which exists on these materials is of sufficient intensity to definitely classify students as potentially inferior or as potential college failures. In order to meet this need for a differential device a three-point classificatory rating scale was constructed.

TABLE IX

COMPARATIVE NORMS FOR THE COMPREHENSIVE RATING SCALE

Measure Used	Scale Value* Scale Score 1 to 19 Rank Order 1	2 20 to 39	3 40 to 59	4 60 to 79	5 80 to 100
A 1a	6	10	14	16	20
1b	6	10	16	20	26
2a	5	7	13	15	17
2b	4	8	9	10	11
2c	4	7	9	11	12
2d	4	7	11	13	17
3a	2	4	6	7	8
3b	0	4	8	10	12
4	12	17	23	28	34
5a	6	10	14	16	18
5b	3	7	12	14	16
6 T.C.	69	121	151	168	186
7 R.R.	10	23	30	33	42
8 Tot.	80	145	190	205	214
B 1	12	22	41	52	67
2	6	13	17	19	23
3 Tot.	32	72	89	103	124
C 1a	3	4	6	7	8
1b	5	10	15	17	20
1c	21	25	30	35	41
1d	11	14	16	17	19
1e	5	18	24	27	33
1f	2	4	6	7	8
1g					
Tot.	63	87	97	102	118
D 1	14	32	48	58	68
2	17	38	57	66	74
3	10	18	26	34	46
4	24	30	52	60	68
5	18	33	45	54	66
6 Tot.	91	156	198	231	273
E	2	21	32	39	55
F 1	76.3	81.2	86.4	93.0	95.6
2	42	61	76	82	93
G 1 (A)	301	337	378	389	415
2 (B)	189	219	238	255	272

*Read the scale from left to right as you look at the page. If any raw score falls exactly upon any rank score, give to the raw score the scale value of the bracket which begins with that raw score. For example, a raw score of 14 for the test A 1a would have a scale value of 3. The same procedure holds true for all rating scale rank orders and values throughout the study, regardless of the ascending or descending order of the scores.

TABLE IX--CONTINUED

Measure Used	Scale Value* 1		2		3		4		5	
	Scale Score 1 to 19		20 to 39		40 to 59		60 to 79		80 to 100	
	Rank Order 1		20		40		60		80	
G 3 (C)	94		108		119		125		134	
4 (D)	124		157		177		191		205	
5 (E)	68		77		85		91		98	
6 (F)	71		82		91		99		110	
7 (G)	75		98		115		124		134	
8 (H)	170		219		251		268		293	
9 (I)	54		62		72		78		84	
10 (J)	114		151		176		185		200	
11 (K)	90		99		114		123		131	
12 (L)	115		133		151		160		172	
13 (M)	138		182		201		209		224	
14 (N)	109		154		167		174		183	
15 (O)	146		201		227		233		240	
16 (P)	101		148		159		164		172	
17 (Q)	79		108		122		127		135	
18 (R)	134		169		193		205		220	
19 (S)	81		106		117		121		127	
H 1a	115		101		90		82		70	
1b	87		74		61		53		44	
1c	64		58		53		49		40	
2a	9.0		7.6		7.1		6.6		5.7	
2b	9.2		7.4		6.5		5.9		5.4	
2c	1.4		1.1		0.8		0.6		0.4	
2d	10		8		6		4		3	
3	15		21		29		33		39	
4a	16		18		21		23		25	
4b	7.2		8.6		10.7		11.4		12.1	
4c	12.2		13.4		14.7		15.3		16.6	
4d	6.2		7.2		8.0		8.6		10.5	
4e	3.8		4.7		5.4		6.0		7.4	
4f	2.1		2.8		3.3		3.7		5.1	
5	12		15		20		22		26	
6	13		16		19		21		24	
7	11		14		19		23		28	
8	7		10		14		16		19	
9	4		6		8		9		11	
10	12		17		22		24		27	
11a	76		85		94		101		112	
11b	48		41		35		30		27	
12a	35		45		56		61		69	
12b	18		14		11		9		8	
12c	75		86		93		96		103	

*Read the scale from left to right as you look at the page. If any raw score falls exactly upon any rank score, give to the raw score the scale value of the bracket which begins with that raw score. For example, a raw score of 14 for the test A 1a would have a scale value of 3. The same procedure holds true for all rating scale rank orders and values throughout the study, regardless of the ascending or descending order of the scores.

TABLE IX--CONTINUED

Measure Used	Scale Value* Scale Score 1 to 19 Rank Order 1	2 20 to 39 20	3 40 to 59 40	4 60 to 79 60	5 80 to 100 80	100
H 13	15	18	20	21	26	30
14	2	3	5	6	7	9
15	6	9	13	15	17	21
16	4	6	8	9	10	12
17	16	25	32	35	40	45
18	3	4	6	7	8	10
19	22	29	34	39	47	56
20	6	8	10	12	13	15
21	94	85	76	71	64	58
I	102	118	126	135	142	149
J	162	181	201	214	223	230
K	110	121	129	145	154	166

*Read the scale from left to right as you look at the page. If any raw score falls exactly upon any rank score, give to the raw score the scale value of the bracket which begins with that raw score. For example, a raw score of 14 for the test A 1a would have a scale value of 3. The same procedure holds true for all rating scale rank orders and values throughout the study, regardless of the ascending or descending order of the scores.

The complete diagnostic instrument was administered individually to the 50 superior, 50 middle, and 50 inferior students described in detail in Chapter II. These student groups were mutually exclusive in terms of their performance on the comprehensive examinations, June 1933.

As a result of the administration of the diagnostic instrument, further norms were derived for each type of diagnostic material by projecting the resulting raw scores of these students upon the comprehensive rating scale and obtaining the mean rating scale scores for the three groups. Four norms were obtained finally for each portion of the diagnostic instrument: (1) the mean score of the superior group; (2) the mean score of the upper 25 students of the middle group; (3) the mean score of the lower 25 students of the middle group; and (4) the mean score of the inferior group. The use of this divided middle group enabled the investigator to set the upper limits for the inferior students and the lower limits for the superior students at definite points on the scale.

The final norms which compose the classificatory rating scale are presented in Table X, together with the probable errors of the differences between the means.

TABLE X
CUMULATIVE NORMS FOR THE CLASSIFICATORY RATING SCALE

Comprehensive Rating Scale Letter	Mean Comprehensive Rating Scale Score				Scale Difference between Superior and Inferior Groups and P.E. $M_1 - M_2$
	Inferior	Lower Middle	Upper Middle	Superior	
A	2.4	3.0	3.9	4.4	2.0 $\pm$.318
B	2.0	3.0	3.7	4.0	2.0 $\pm$.295
Cum. Total . . .	4.4	6.0	7.6	8.4	
C	2.0	2.7	3.7	4.1	2.1 $\pm$.327
Cum. Total . . .	6.4	8.7	11.3	12.5	
D	1.8	2.5	2.8	4.0	2.2 $\pm$.326
Cum. Total . . .	8.2	11.2	14.1	16.5	
E	2.0	3.0	3.0	3.0	1.0 $\pm$.193
Cum. Total . . .	10.2	14.2	17.1	19.5	
F	2.0	3.0	3.5	4.5	2.5 $\pm$.315
Cum. Total . . .	12.2	17.2	20.6	24.0	
G	2.1	2.8	3.7	4.4	2.3 $\pm$.297
Cum. Total . . .	14.3	20.0	24.3	28.4	
H	2.6	3.0	3.8	3.9	1.3 $\pm$.201
Cum. Total . . .	16.9	23.0	28.1	32.3	
I	1.0	2.0	3.0	4.0	3.0 $\pm$.386
Cum. Total . . .	17.9	25.0	31.1	36.3	
J	2.0	3.0	3.0	4.0	2.0 $\pm$.261
Cum. Total . . .	19.9	28.0	34.1	40.3	
K	3.0	3.0	3.0	3.0	0.0 $\pm$.204
Cum. Total . . .	22.9	31.0	37.1	43.3	
Mean	2.1*	2.8	3.4	3.9	1.9 $\pm$.285**
Total Excluding E, H, J and K	13.3	19.0	24.3	29.4	
Mean Excluding E, H, J and K	1.9*	2.7	3.5	4.2	2.3 $\pm$.311**

*Means may be averaged when all of the N's are equal (Holzinger, *op. cit.*, p. 84).

**These P.E. values are not averages, but are computed from the original data.

An analysis of the data presented in this table shows that there is considerable variability in the differential value

of the various diagnostic materials. The order of descending value as indicated by the total scale difference between the superior and inferior students is as follows: (1) ratings on scholastic traits; (2) a combination of high-school average and high-school rank; (3) the reading and study check lists; (4) intelligence or scholastic aptitude; (5) the English placement test; (6) the Iowa Reading Test; (7) the Minnesota Reading Examination; (8) the ratings on personality traits and characteristics; (9) the informal individual tests; (10) the Personality Schedule; and (11) the ratings on personal appearance and hygiene.

Under the more favorable combination of the diagnostic materials, obtained by excluding sections E, H, J, and K, a differential area of $2.3 \pm .311$ comprising over half of the entire five-point scale lies between the mean performance of the superior and inferior groups. The significance of the differences on each type of diagnostic material is shown statistically by the probable error of the difference between the means. The general rule in regard to these differences is that they are not significant unless they are four times the probable error.¹ The differences, as reported in Table X, are significant therefore for all sections of the scale, excepting section K.

In connection with the construction of this scale it was necessary to sacrifice some of the classificatory or differential value of the diagnostic materials in order to obtain a complete individual diagnosis. For example, section K, which consists of ratings on personal appearance and hygiene, has no differential value for the reason that the superior students were not superior on all of the items which compose that measure. The differential value of Section K was canceled out by this tendency. The major purpose of the study, however, was to obtain a complete diagnosis of individual students as well as to obtain a general differentiation between superior and inferior students. The elimination of all items or measures which did not possess great differential value would have reduced the diagnostic instrument to the status of a purely differential device, thus decreasing the diagnostic value to a minimum. Since there is a difference of $2.3 \pm .311$, or over half of the five-point scale range, between the mean performance of superior and inferior students on the more favorable

¹Holzinger, *op. cit.*, p. 237.

combination of the diagnostic materials, the classificatory function of the instrument is adequately fulfilled.

Although the original construction of the scale is rather complex, its subsequent use by administrative officers is relatively simple. After the administration of the desired portion of the diagnostic materials, a student's raw scores may be projected upon the comprehensive rating scale to reduce them to a comparable basis. If it is desired to determine whether the quality of these scores is such that they will contribute to the student's potential scholastic inferiority or superiority, they may be projected, in turn, upon the classificatory rating scale and given the proper rating. For example, since the scores of the divided middle group on section A, the Iowa Reading Test, are 3.0 and 3.9, any score below 3.0 would be classified as inferior and any score above 3.9 as superior. The relative degree of this superiority or inferiority may be determined by comparing the score with the mean of the group under which it falls. For example, a score of 2.0 on Section A indicates that the score should be classified as inferior. Since the score falls below the mean of the inferior group, however, it should be classified as low inferior. The scale was not designed, however, to obtain these finer degrees of classification, but merely to obtain a simple yet adequate means of differentiating between the extremes, that is, the superior and inferior students. The inferior students may be identified thus for purposes of remediation.

Two purposes are achieved, therefore, by the administration of the diagnostic instrument: (1) a comprehensive diagnosis is obtained of the abilities or deficiencies of students on all of the diagnostic materials; and (2) a general classification is obtained for students which identifies them in terms of the superior and inferior categories. The diagnostic materials and statistical devices described in the foregoing sections of the chapter furnish the basic instruments for attaining these two objectives. The actual use of the scales and of the diagnostic instrument is illustrated further in the diagnostic and remedial experiment reported in Chapter V.

The diagnostic materials described up to this point comprise the portion of the instrument which was partially standardized in the study. An additional diagnostic measure, which is essentially unstandardized, was developed by the investigator to

explore those areas or factors of experience, training, or environment which might possibly condition or affect either scholastic success or reading and study abilities, but which could not be measured by tests or rated on scales, adequately. This measure is described in the succeeding section of the chapter.

The Unstandardized Exploratory Personal and Family Data Sheet

The major function of the standardized portion of the diagnostic instrument is to arrive at a knowledge of the present status of a student's reading and study abilities or deficiencies. No diagnosis would be adequate, however, which did not attempt to determine the causal or concomitant factors of experience, training or environment which may have led to the development of the specific abilities or deficiencies. This is especially important in determining the proper remedial procedures to overcome the deficiencies. If the remedial worker attempts to correct a deficiency without attempting to remove, condition, or correct the cause of that deficiency or the situation in which it occurs, if possible, a significant portion of the remedial work remains unfinished.

The determination of these causal or conditioning factors can not be accomplished through the use of formalized tests, rating scales, or similar measuring devices. Probably the best means of exploring this phase of a diagnosis is through the ordinary interview method. In order to inject some element of control into the interview technique and to focus its content upon certain specific objectives, some definite plan of attack or exploration should be used by the interviewer.

To meet this need the investigator made a critical analysis of many types of information blanks, personnel schedules, data sheets, family and individual case histories, and similar materials to determine the factors which seem to be most closely related to reading and study deficiency or scholastic failure. As a result of this analysis, a set of personal and family data sheets was developed which serves to control the interview by centering it upon or focusing it around certain key items of information. These data sheets are composed of 224 items which attempt to explore those factors which may have contributed, genetically to the development of good and poor reading and study

abilities or to scholastic success. Only those factors or items were selected which had been shown to be contributory to deficiency in the case histories reported in previous studies.¹

Since the data sheets were designed to be used primarily by the remedial worker, rather than by the student, a check list form of response was adopted to save time during the interview. For example, the item number 78 is as follows:

78. What is your economic status in college?

1. Totally supported by parents, relatives, or income.
2. Partially self-supporting.
3. Totally self-supporting.

It is apparent that the proper response can be checked very readily by the interviewer after he has asked the subject leading questions concerning the specific item or area which is being explored. This check list form of response results not only in a saving of time but reduces all of the responses to comparable terms.

Items of this general type were included which cover, genetically, the period from birth to college entrance. They attempt to explore factors of general experience, education, play life, reading and study activities, vacation activities, vocational and avocational interests, physical efficiency and deficiency, family background, and many others. The knowledge gained from the use of the data sheets gives a rather complete picture of the setting or situation in which the specific reading or study deficiency has developed or exists. This informal exploratory device rounds out or supplements the standardized portion of the diagnostic instrument and aids in laying an adequate basis for the institution of the proper remedial procedures.

The data sheets were administered by the investigator to the 50 superior and 50 inferior students to determine the items which might be of some differential value, or which seemed to have led, genetically, to the development of good or poor reading and study abilities as well as scholastic success. As a result of this administration, it was found that approximately 119 of the 224 items bore some significant relationship, either causal or concomitant, to the development of these abilities. A summary of these differences is reported in the paragraphs that follow.

¹Gray, op. cit.; Cowley, op. cit.

The average age of the superior students at college entrance was 18.3 years, as against 19.2 for the inferior students. The percentage of women in the superior group was 20 as against 42 in the inferior group. More of the superior students attended nursery school and kindergarten. The superior students spent fewer number of years in elementary school and high school, the number being 7.2 and 3.8, respectively, as against 8.8 and 4.2 for the inferior students. The superior students entered high school at an average age of 13.3 and the inferior students at an average age of 14.2 years.

The average superior student was born in a city of 55,000 population and the average inferior student in a city of 19,000 population. The superior students moved a lesser number of times since birth, the number being 1.2 for the superior group as against 1.7 for the inferior group. Eighty-five per cent of the superior group and 50 per cent of the inferior group lived in Chicago at the time of the administration of the data sheets. The superior students attended larger high schools, and went to elementary schools in larger cities.

Twenty-seven of the superior students were able to read when they entered the elementary school as against 14 of the inferior students. More of the superior students were taught to read widely rather than analytically in high-school English and social science courses. More of the superior students were encouraged to read widely at home and at school, as children. More of the superior students were liked by their teachers in elementary school, but fewer were liked by their teachers in high school.

More of the superior students had opportunity classes in elementary school and high school and more of them were actually members of such opportunity classes as reported by the subjects. The average mark earned by the superior students in elementary school was 95.1 and by the inferior students, 91.2. More of the superior students had been taught by the unit method and had used the project, laboratory, contract, and other of the newer types of learning procedures.

The superior students failed fewer courses and were placed on probation less often for poor scholastic work in elementary school and high school. The superior group repeated a total of one grade in elementary school as against seven for the inferior group.

More of the students in the superior group expected to take graduate or professional training after completing their college work; 36 superior students reporting this intention, as against 31 of the inferior students. The superior students missed fewer weeks of school in the elementary grades. More academic encouragement was received by the superior students in elementary school, high school, and college; 49 of the superior students reporting this item favorably as against 29 of the inferior students. Both groups had read considerably as children and continue to read extensively at present. Fewer superior students reported difficulty in covering assignments in elementary school, high school, and college. Fewer superior students had encountered specific reading or study difficulties at any time. More superior students lived at home throughout their elementary school, high school, and college careers. Thirty-three of the superior students had chosen a profession such as law or medicine, 18 reporting this choice as against 13 of the inferior students. Similarly, 24 of the superior group and only 19 of the inferior group were seeking the doctorate for their final degree. The superior group spoke and read fewer languages, probably due to the fact that their racial origin was more homogeneous than that of the inferior group.

The superior students participated in more scholastic or academic extra-curricular activities in high school, and won more academic honors, prizes, and medals. Fewer athletic activities were engaged in and fewer trophies were won by the superior students. Twenty-two of the superior students held scholarships at the University as against nine of the inferior students. Little difference was apparent in the number of hours spent in study on week-days in college, the averages being 3.8 and 3.7 for the superior and inferior groups, respectively. On Saturdays, however, the superior group spent 2.1 hours in study as against 1.8 for the inferior group, and on Sundays the superior group spent 3.4 hours in study as against 3.1 for the inferior group.

The superior group spent more time commuting to college, reporting an average of 22 minutes per day as against 15 minutes per day for the inferior group, probably due to the fact that more of the superior group lived at home. Of the superior group, 21 studied best in the later part of the evening as against 16 of the inferior group. The superior group spent an average of 39.6

hours per month in reading along lines of special interest or hobbies as against 24.7 for the superior group. The superior group spent only .9 hours per day in part-time employment as against 2.6 for the inferior group. Forty-one of the superior group and 30 of the inferior group were supported totally by parents or by income other than that derived from part-time employment. The school, personal, and home libraries of the superior students were larger than those of the inferior students. Thirty-nine of the superior students used the libraries voluntarily as against 28 of the inferior group. A greater number of minutes per day was spent in reading in libraries and at home by the superior students. The superior students read more non-fiction and good fiction than the inferior students. The high school work of 41 of the superior and 33 of the inferior students was well-rounded academically.

Forty-seven of the superior students and 10 of the inferior students felt assured of future success in college work, while 48 of the superior students and 41 of the inferior students had felt well prepared and assured of success when they came to college. The superior students listened to the radio less, attended the "movies" less, and had fewer "dates" and "bull-sessions". When the superior students listened to the radio they preferred more educational and current events discussions and classical music, and less "jazz" music than the inferior students. Twenty-two of the superior students and 10 of the inferior students read the editorial page of the daily newspapers. Nineteen of the superior students and 16 of the inferior students read the newspapers for general information. Forty-two of the superior group and 38 of the inferior group stated that they would read more if they had more time. The average superior student retired at 11:10 o'clock and the average inferior student retired at 11:30 o'clock. The superior group read an average of 12.4 magazines per month as against 9.6 for the inferior group. The superior group read more non-fiction and good fiction magazines than the inferior group. More books were read thoroughly or completely by the superior group, since they read an average of 296 books per year as against 249 by the inferior group. The superior students read more scientific and non-fiction books than the inferior students. Thirty-six of the superior students skimmed fiction materials, in general, as against 29 of the inferior students.

Thirty-four of the superior students read non-fiction thoroughly as against 28 of the inferior students. The superior students commenced to read books, magazines, and newspapers at an earlier age than the inferior students. Thirty-two of the superior students and 25 of the inferior students preferred long stories and articles. More of the superior students actually enjoyed non-fiction, and 36 of the superior students preferred non-fiction as against 24 of the inferior students. More of the superior students had read considerably during convalescences.

Fewer of the superior students had been disciplinary problems at home or at school. More of the parents of the superior students had used praise and fewer had used corporal punishment as means of controlling their children. Twelve of the superior students and eight of the inferior students reported that they played alone as a child. More of the superior students were of the "visual-minded" type and fewer of the superior students were of the "aural-minded" type. More of the superior students reported that they were getting what they came to college for and that they were satisfied with themselves, in general. Fewer of the superior students had many friends and fewer of them made friends easily than the inferior students. More of the superior students had a tendency to hold the friends they did make than the inferior students.

More of the superior students had a definite philosophy of life, 37 reporting this item as against 29 of the inferior students. Similarly, more of the superior students were liberal on moral issues. Nineteen of the superior students and 12 of the inferior students stated that they were very temperamental, while 35 of the superior group and 27 of the inferior group reported that they were very imaginative. More of the superior students found the college lectures understandable and the readings reasonable in difficulty. More of the superior students read the optional readings. The superior students found the readings reasonable in length and felt that the amount of work required under the New Plan was reasonable. The comprehensive examinations were considered reasonable by more of the superior students. More of the superior group felt that their mastery of the New Plan course materials was satisfactory. More of the superior students found the writing of papers helpful, but fewer of them found the syllabi satisfactory or the advisers helpful. Fewer of the superior stu-

dents had difficulty in preparing for the comprehensive examinations and more of them stated that they were stimulated to do their best work under the "New Plan."

The parents of superior students were an average of 4.5 years younger than the parents of inferior students. The superior students had fewer parents dead, divorced, or separated. The ancestry of the superior students and of their parents was more homogeneous, being 80 per cent Nordic as against 66 per cent for the inferior students. In the homes of 48 of the superior students and 43 of the inferior students the English language was used in ordinary conversation. More members of the families of superior students were born in the United States and had maintained longer residence in this country than had similar members of the families of inferior students. More of the parents of superior students had attended high school, college and university, and were engaged in teaching and professional fields. The average yearly income of the fathers of superior students was \$2,500 as against \$2,100 for the fathers of inferior students. More of the parents of superior students had special professional or academic interests, hobbies, or talents. More of the parents of superior students read for recreation. The parents of superior students were born and were living in larger population centers. More parents of superior students were active in civic, literary, or artistic clubs. Superior students had fewer siblings and relatives, although they had a greater percentage of siblings and relatives who had attended and graduated from high school, college, and university.

Superior students reported that 1.4 daily newspapers were received in the home as against 1.1 for the inferior students. The superior students reported that 4.2 magazines were received in their homes per month as against 3.9 for the inferior students. The superior students reported that they received fewer fiction and more non-fiction magazines in the home. More superior students reported that their parents and siblings read the editorial page of the daily newspaper than the parents and siblings of inferior students. The parents and siblings of superior students listened to more educational and current events discussions, and to classical music on the radio than did the parents and siblings of inferior students. Finally, the parents and male siblings of

superior students did more voluntary reading per day than those of inferior students.

The foregoing summary of differences indicates that the items of the personal and family data sheets have considerable differential value. This is especially true in case any number of the responses to these significant items form favorable or unfavorable combinations similar to those which favor either the superior or inferior students. If such combinations should occur in addition to similar favorable or unfavorable scores on the comprehensive or classificatory rating scales, the more objective differential criteria would be strengthened proportionately in individual cases.

Use of the Diagnostic Materials

The diagnostic materials described in this chapter compose the comprehensive instrument developed in the study. The general use to which these materials may be put is twofold: (1) to classify students as superior or inferior as a means of identifying the students who are most in need of remediation; and (2) to obtain a complete diagnosis of the general abilities or specific deficiencies of individual students.

The use of the instrument as a means of classification involves the following steps: (1) the administration of the diagnostic materials; (2) the projection of the resulting raw scores upon the comprehensive rating scale to reduce all of the scores to a comparable basis; and (3) the final projection of the rating scale scores upon the classificatory scale to determine whether a general classification of superior or inferior may be given. Through this use of the diagnostic technique students who are most in need of remedial training may be identified readily.

After the preliminary identification of those students who are most in need of remediation, the use of the instrument as a means of individual diagnosis becomes paramount. The same scores derived and used in connection with the classificatory function of the instrument are used further in the diagnostic function.

In determining the general areas of deficiency either the comprehensive or the classificatory scales may be used. When using the comprehensive rating scale as the instrument of diagnosis, the scores of 1 and 2 indicate relative degrees of defi-

ciency, a score of 3 indicates average ability, and scores of 4 and 5 indicate relatively superior ability on each type of diagnostic material. It is possible to use this scale in the manner suggested, owing to the fact that it was constructed from the scores of typical samples of the parent population on these materials. The distribution of scores, as represented on the five-point scale, is typically representative therefore of the total distribution of scores made by the parent population and gives a quality scale ranging from the highest to the lowest.

Whether the degree of deficiency that is indicated on the comprehensive scale is sufficient to classify the student as inferior to the extent of being a potential college failure can not be determined, however, without the use of the classificatory scale. Although a student may have been classified previously as inferior in terms of his final score on the entire technique, there may be certain portions of the scale upon which his rating may not be markedly inferior. After the preliminary use of the classificatory scale to identify students who have general deficiency, it may be used further to identify the specific portions of the instrument in which the greatest deficiency is present. This is accomplished by comparing the student's rating on each type of diagnostic material with the proper norm of the classificatory scale and determining whether a classification of inferior may be given to the student on that particular area. Through this use of the scale, a preliminary identification may be made of the general areas of deficiency, thus laying the foundation for the more specific diagnosis which follows.

The determination of specific remedial procedures needed to overcome deficiencies demands, however, a more specific type of diagnosis than is obtained through the use of the rating scales described in this chapter. After the identification of the general areas of deficiency, a more detailed diagnosis may be obtained by analyzing the actual responses made by the student on the individual items of the tests, sub-tests, and reading and study check lists. The wrong responses on the tests and sub-tests may be identified readily by the use of the proper scoring keys. On the reading and study check lists the specific items of deficiency may be determined by identifying all of the items to which the student made the responses 1 and 2, or never and seldom. Since the items in these lists are all stated in positive form

and are good procedures, failure to use these procedures may be taken as a sign of deficiency.

The actual identification of the specific items of deficiency may be supplemented further by an analysis of the responses made by the students to the items of the personal and family data sheets, as well as of the Personality Schedule, to determine the causal or concomitant factors of personality, training, or environment which might have contributed to the development of deficiency. By attempting to condition the modifiable factors which are unfavorable to reading and study efficiency or scholastic success, the remedial worker lays the foundation for the direct or indirect improvement of the scholastic efficiency of the individual student. Even if he does not find specific contributory factors to the existing deficiency, the investigator will obtain a general understanding of the individual case which will prove of value in any type of remedial training that may be undertaken.

The foregoing discussion indicates the general use to which the diagnostic measures may be put, either for purposes of classification or of diagnosis. An actual illustration of the use of the instrument is given in Chapter V which describes the diagnostic and remedial experiment. Similarly, many suggestions for the further use and study of the instrument beyond that made in the present investigation are given in the final chapter, which considers the findings and interpretations of the study.

Since norms are available for each portion of the diagnostic instrument, as well as for the entire instrument, the program of diagnosis may be made as flexible and as comprehensive as is necessary in a particular administrative situation. The actual procedure used in this study may be followed in detail or may be deviated from to whatever degree the remedial worker deems desirable, since the instrument was designed to be used as an informal rather than a formal type of diagnostic measure. The procedure followed by the investigator in organizing and conducting the diagnostic and remedial experiment in succeeding chapters is intended to be suggestive rather than mandatory.

CHAPTER IV

THE ORGANIZATION OF THE DIAGNOSTIC AND REMEDIAL EXPERIMENT

Problem of the Chapter

It was necessary to organize an adequately controlled experiment in order to apply the diagnostic instrument and related remedial procedures to potential college failures to determine their administrative value. The organization of this experiment involved four major problems or steps: (1) the preliminary identification of students who could be classed as potential college failures; (2) the administration of the diagnostic instrument to the potential college failures to determine the students who had reading and study deficiencies; (3) the setting up of adequate experimental controls; and (4) the introduction of the experimental variable, which, in this study, was the remediation of reading and study deficiencies. This chapter is concerned with the discussion of the first three problems.

Derivation and Statement of the Methods of Predicting Potential College Failures

It was impossible for the investigator to administer the complete battery of diagnostic materials, individually, to the entire class of 700 freshmen entering the University of Chicago in the autumn quarter, 1933. It was necessary, therefore, to make a preliminary study to identify students who could be classed as potential college failures on the basis of the tests and data already available in the various offices of the University.

In order to identify the potential college failures, two preliminary predictive methods were used. One method, derived by Reitz¹ under "New Plan" conditions, was based upon a comprehen-

¹William Reitz, "Forecasting Success of Students under the New Plan at the University of Chicago." Unpublished report to the Dean of Students and University Examiner, University of Chicago, May, 1933.

sive study of the zero order and multiple correlations that existed between marks on the comprehensive examinations in the four general survey courses of the College and three predictive factors, namely, intelligence, average of high-school marks, and rank in high-school graduating class. From a study of the relative forecasting value of these three factors, regression and multiple forecasting formulas were derived by Reitz for the prediction of success in the four comprehensive examinations. The statistical procedures used to derive these formulas are presented in detail in Reitz's study and will not be discussed here.

A second method of predicting potential college failure was based upon the percentage of various marks actually earned during the two previous years, 1931-32 and 1932-33, by students who fell below certain critical scores on four predictive factors. The factor of reading comprehension was added to the three measures used by Reitz, since it was necessary to select subjects for the remedial experiment who were deficient in general reading ability as well as potential college failures.

The general predictive value of intelligence, high-school average, and high-school rank was determined from the relative size of the coefficients derived when these factors were correlated with marks on the comprehensive examinations. The size of these coefficients is shown in Table XI.

TABLE XI
INTER-CORRELATIONS BETWEEN ALL PREDICTIVE VARIABLES*

Variables	1	2	3	4	5	6	7
1. Rank in high-school	...	.72	.27	.45	.34	.52	.46
2. Average high-school mark. .	.72	...	.34	.52	.46	.49	.52
3. Scholastic aptitude	.27	.34	...	.49	.64	.58	.52
4. Mark in Biological Science. .	.46	.52	.49	...	.73	.86	.77
5. Mark in Humanities.	.34	.46	.64	.73	...	.72	.84
6. Mark in Physical Science. .	.52	.49	.58	.86	.72	...	.71
7. Mark in Social Science. . .	.45	.52	.52	.77	.84	.71	...

*Adapted from Reitz, op. cit., p. 5.

The average of the coefficients, when the three predictive factors were correlated with marks on the comprehensive examinations, as computed from the data in Table XI, was .56 for intelli-

gence or scholastic aptitude, .50 for average of high-school marks, and .43 for rank in high-school graduating class. The descending order of the predictive value of these three factors, therefore, is intelligence, high-school average, and high-school rank.

Using intelligence, the factor of highest predictive value, as the first basis of selection, tabulations were made of the various marks actually earned on the comprehensive examinations during the two preceding years, 1931-32 and 1932-33, by students ranking in the lower percentiles or having low scores on this measure. By a process of progressive elimination, the seventh percentile in intelligence was finally determined to be the critical point at which a satisfactory prediction of F or failure could be made in the greatest number of individual cases. By a similar process of tabulation of marks and progressive elimination, the critical points for a satisfactory prediction of F were determined for the three additional measures. These combined critical points were found to be a high-school average of 85 or below, the 16th percentile or below in reading comprehension, and a percentile rank of 86 or below in the high-school graduating class. At any point above these combined predictive critical points the dispersion of marks increased rapidly; at or below these points the homogeneity of the marks increased.

The data presented in Table XII show the predictive value of the four combined factors in terms of the number and percentage of various marks actually earned on the comprehensive examinations by students of the two previous New Plan classes, 1931-32 and 1932-33, who fell in the four critical categories.

The entries in this table show that 78.2 per cent of the total number of marks earned by students who fell in the four categories were F or of failing quality. To that degree, therefore, students who are in the four combined predictive categories may be classed as potential college failures.

This distribution of marks was used further as a basis of comparison for evaluating the final scholastic results obtained in the diagnostic and remedial experiment. The comparison was made in terms of the percentage of failing marks actually earned on the comprehensive examinations by the experimental and control groups. These two groups were selected on the basis of the same

TABLE XII
DISTRIBUTION OF MARKS EARNED ON THE COMPREHENSIVE EXAMINATIONS BY SELECTED STUDENTS OF THE CLASSES OF 1931-32 AND 1932-33 WHO FELL BELOW THE CRITICAL PREDICTIVE POINTS IN INTELLIGENCE, READING, HIGH-SCHOOL AVERAGE, AND HIGH-SCHOOL RANK

Mark	Class of 1931-32 (23)		Class of 1932-33 (23)		Combined Classes	
	No.	Per Cent	No.	Per Cent	No.	Per Cent
A						
B						
C	2	3.5	4	7.5	6	5.4
D	8	14.0	10	18.9	18	16.4
F	47	82.5	39	73.6	86	78.2
Total	57	100.0	53	100.0	110	100.0

combined predictive measures used in isolating the class groups shown in Table XII.

After the preliminary determination of the critical predictive points had been made by the method discussed above, the multiple forecasting formulas constructed by Reitz were used to obtain a refined statistical prediction of potential college failure. Three of the predictive criteria were substituted in the forecasting formulas: (1) the seventh percentile in intelligence; (2) a high-school average of 85; and (3) a percentile rank of 86 in the high-school graduating class. By substituting these values in each forecasting formula and projecting the resulting predicted value upon the proper scale, as indicated by Reitz, predictions of F or failure were obtained for all of the subjects used in the experiment on all four of the comprehensive examinations.

The formulas used to obtain these predictions and the procedure followed in using them are shown in detail in the report by Reitz.¹ For illustrative purposes one example of the procedure is presented at this point. The formula for the prediction of success in the comprehensive examination of the Biological Science

¹Op. cit.

General Course is shown below with the necessary substitutions and legend.¹

Score Form. Three Variables. (r = .63)

$$B_{est.n.} = .096960H_p + .717571HM + .121364SA - 62.041786 \text{ (Formula)}$$

$$B_{est.n.} = 8.338560 + 60.992135 + .728184 - 62.041786 \text{ (Substitutions)}$$

$$B_{est.n.} = 8.02$$

By projecting the derived predicted value upon the proper scale it was found that a mark of F or failure was predicted for all students at or below the combined predictive categories, as follows:

	Scale	
	7.72	F
Predicted Value 8.02	-----	
	10.75	D

The predicted values obtained through the use of the four predictive formulas were as follows: Biological Science, 8.02; Humanities, 5.18; Physical Science, 10.40; Social Science, 8.97. Each of these values gave a prediction of F or failure when projected upon the proper scale. The average predicted value on the four examinations was 8.14, which gave a general prediction of F or failure on all four comprehensive examinations when projected upon the proper scale, as follows:

	Scale	
	7.23	F
Predicted Value 8.14	-----	
	10.43	D

Since the raw scores and percentile ranks substituted in the forecasting formulas were critical or maximal in nature, the statistical predictions derived through their use were, likewise, critical or maximal. Predictions derived at these critical points hold true, therefore, for all values below these points. In se-

¹Legend:--

- B Mark in the Comprehensive Examination in the Biological Sciences
- p Percentile form
- est.n. Estimated variable, numerical form
- HM Average of high-school marks
- RH Rank in high-school graduating class
- SA Percentile rank on the American Council Psychological Examination

lecting the subjects used in the experiment the combined values of the predictive measures were held, in every case, at or below the maximal points. It is evident from the discussion of predictive methods presented in this section of the study that the individuals selected for the diagnostic and remedial experiment could be classified as potential college failures on the basis of the four critical predictive measures.

Preliminary Identification of the Subjects Used in the Experiment

The foregoing method of predicting potential college failure was applied to the entering freshman class in the autumn of 1933 in order to identify the subjects to be used in the diagnostic and remedial experiment. The regular placement testing program was administered by the Board of Examinations during freshman week. Immediately after the administration and scoring of these tests by the Board, all individual students were identified for preliminary investigation who fell at or below the critical points on the four predictive measures. As reported earlier, these critical points were the seventh percentile rank in intelligence, a high-school average mark of 85 or below, a rank of 86 or below in high-school graduating class, and a percentile rank of 16 or below in reading comprehension. The actual scores and ranks of each student are reported later in Table XV. The percentile ranks were obtained from the Board of Examinations while the data used to compute the high-school averages and ranks in high-school graduating classes were obtained from the application for admission blanks in the office of the Dean of Students in the College.

A preliminary group of 39 students who could be classed as potential college failures was identified in this manner for further investigation. The application of the next qualification, that of registration in college courses which would permit of the selection of matched pairs, eliminated 13 of the group of 39 students. The battery of diagnostic materials in reading and study was then administered to the remaining 26 subjects to determine their abilities and deficiencies.

Administration of the Comprehensive Diagnostic Instrument to the Subjects Used in the Experiment

When the group of potential college failures had been identified by means of the general predictive methods described, it was necessary to obtain a further diagnosis and classification of the reading and study abilities of these subjects as a basis for the subsequent remedial procedures.

With the permission and cooperation of the Dean of Students the 26 potential college failures were summoned to the advisory staff offices to meet the investigator, individually. Their cooperation was asked in an experiment to determine the value of the diagnostic instrument and related remedial procedures developed in this study. The diagnostic materials were administered to the subjects, individually, by the investigator during the period beginning October 23 and ending November 4, 1933.

Three results were secured from the diagnostic data thus obtained: (1) a general summation of the performance of each subject on the diagnostic materials was obtained for purposes of classification on the classificatory rating scale and for further matching of the subjects; (2) a complete diagnosis of the reading and study deficiencies of each subject was obtained; and (3) the specific diagnosis of deficiencies laid the basis for the subsequent remedial procedures. After the administration of the diagnostic materials, three additional subjects were eliminated from the experiment as not having a marked degree of deficiency in reading and study.

The scores derived from the administration of the diagnostic materials were tabulated for each subject and projected upon the comprehensive five-point rating scale in order to obtain a final score which could be used in the further refinement of the experimental matching procedure. Scores were obtained for each division of the scale; cumulative summations were made of the scores on all divisions of the scale; and a final score was derived which furnished the basis for the general classification of these subjects into the category, inferior, when projected upon the classificatory scale. Through this procedure each test, sub-test, check list, and informal personal rating was made to contribute its proportionate share to the final score on the diagnostic instrument.

The rating scale scores made by the subjects who were matched later into experimental and control pairs are shown in Tables XIII and XIV.

The cumulative or summated rating scale scores shown in these tables gave a general classification of inferior when they were projected upon the classificatory scale shown in Table X. For example, the range of the middle classification on Section A of the scale, the Iowa Reading Test, is from 3.0 to 3.9. Any score below 3.0 on this section of the scale may be classed, therefore, as inferior. The highest score earned by any of the 23 subjects was 2.4 and the lowest score was 1.5. It is obvious that these subjects should be classed as inferior on this portion of the diagnostic materials.

The range of the middle classification on the total or final score, as shown in Table X, is from 2.7 to 3.5. The highest final score earned by the 23 subjects was 2.1 and the lowest was 1.4. Since all final scores below 2.7 should be classified as inferior, it is apparent that all of the subjects used in the experiment belong in this general category. The comparative norms for the inferior section of the classificatory scale were derived from the scores of the 50 students who received either three F's or two F's and one D on the college comprehensive examinations. It was possible, therefore, to make a supplementary prediction of potential failure for these subjects owing to their deficiencies on the diagnostic materials.

The Matching or Pairing of the Experimental and Control Subjects and Groups

One of the major problems involved in the organization of the diagnostic and remedial experiment was the setting up of adequate controls to hold constant the effect of variables other than the remedial procedures conducted by the investigator.

The matched pair and parallel group method of experimentation was used in this portion of the study because of its simplicity and scientific adequacy. These methods were designed for use in experiments in which it is necessary to hold constant or to rule out the operation of factors or variables other than those introduced into the experiment by the investigator. According to this method, each subject in the experimental group was matched or paired with a similar subject in the control group who

TABLE XIII
RATING SCALE SCORES OF THE EXPERIMENTAL SUBJECTS ON THE FORMAL TESTS,
INFORMAL TESTS, CHECK LISTS AND INFORMAL PERSONAL RATINGS

Rating Scale Letter	Case Numbers of Experimental Individuals												
	Experimental Group A										Exper. Group B		
	1	2	3	4	5	6	7	8	9	10	1	2	3
A (Iowa)	1.9	1.5	2.3	2.1	1.7	2.4	2.1	1.9	1.8	2.0	1.8	1.6	1.9
B (Minnesota)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Cumulative Total	2.9	2.5	3.3	3.1	2.7	3.4	3.1	2.9	2.8	3.0	2.8	2.6	2.9
C (English Placement)	1.3	1.7	2.4	1.6	2.4	3.0	2.3	1.6	1.3	2.3	1.4	1.6	1.3
Cumulative Total	4.2	4.2	5.7	4.7	5.1	6.4	5.4	4.5	4.1	5.3	4.2	4.2	4.2
D (Psychological)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Cumulative Total	5.2	5.2	6.7	5.7	6.1	7.4	6.4	5.5	5.1	6.3	5.2	5.2	5.2
E (Personality)	1.0	4.0	2.0	4.0	1.0	2.0	5.0	2.0	2.0	5.0	1.0	3.0	1.0
Cumulative Total	6.2	9.2	8.7	9.7	7.1	9.4	11.4	7.5	7.1	11.3	6.2	8.2	6.2
F (H.S. Rank and Av.)	2.0	2.5	3.0	3.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	1.5
Cumulative Total	8.2	11.7	11.7	12.7	10.1	11.4	13.4	9.5	9.1	13.3	8.2	11.2	7.7
G (Check Lists)	1.5	1.7	2.1	2.2	2.5	1.9	2.0	2.0	2.6	1.7	2.2	1.5	2.7
Cumulative Total	9.7	13.4	13.8	14.9	12.6	13.3	15.4	11.5	11.7	15.0	10.4	12.7	10.4
H (Informal Tests)	2.1	2.3	2.4	2.5	2.5	2.7	2.5	2.3	2.5	2.7	2.5	2.3	2.6
Cumulative Total	11.8	15.7	16.2	17.4	15.1	16.0	17.9	13.8	14.2	17.7	12.9	15.0	13.0
I (Scholastic)	1.0	1.0	1.0	2.0	3.0	1.0	1.0	1.0	2.0	1.0	2.0	1.0	1.0
Cumulative Total	12.8	16.7	17.2	19.4	18.1	17.0	18.9	14.8	16.2	18.7	14.9	16.0	14.0
J (Personal Traits)	2.0	2.0	2.0	2.0	3.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
Cumulative Total	14.8	18.7	19.2	21.4	21.1	18.0	20.9	16.8	18.2	20.7	17.9	18.0	16.0
K (Personal Appearance)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	2.0
Cumulative Total	17.8	21.7	22.2	24.4	24.1	21.0	23.9	19.8	21.2	23.7	20.9	21.0	18.0

TABLE XIII--CONTINUED

Rating Scale Letter	Case Numbers of Experimental Individuals												
	Experimental Group A										Exper. Group B		
	1	2	3	4	5	6	7	8	9	10	1	2	3
Mean (All Items). . .	1.6	2.0	2.0	2.2	2.2	1.9	2.2	1.8	1.9	2.2	1.9	1.9	1.6
Total Excluding E, H, J, and K	9.7	10.4	12.8	12.9	14.6	12.3	11.4	10.5	11.7	11.0	11.4	10.7	10.4
Mean Excluding E, H, J, and K	1.4	1.5	1.8	1.9	2.1	1.8	1.6	1.5	1.7	1.6	1.6	1.5	1.5

TABLE XIV

RATING SCALE SCORES OF THE CONTROL SUBJECTS ON THE FORMAL TESTS, INFORMAL TESTS, CHECK LISTS AND INFORMAL PERSONAL RATINGS

Rating Scale Letter	Case Numbers of Control Individuals									
	1	2	3	4	5	6	7	8	9	10
A (Iowa).	1.9	1.6	1.9	2.4	2.1	2.4	2.4	2.1	1.8	2.2
B (Minnesota).	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Cumulative Total.	2.9	2.6	2.9	3.4	3.1	3.4	3.4	3.1	2.8	3.2
C (English Placement).	1.7	1.9	2.1	1.7	2.4	2.0	1.4	1.9	2.3	2.6
Cumulative Total.	4.6	4.5	5.0	5.1	5.5	5.4	4.8	5.0	5.1	5.8
D (Psychological).	1.0	1.0	1.2	1.2	1.0	1.0	1.0	1.0	1.0	1.2
Cumulative Total.	5.6	5.5	6.2	6.3	6.5	6.4	5.8	6.0	6.1	7.0
E (Personality Schedule).	2.0	2.0	2.0	3.0	5.0	1.0	1.0	2.0	1.0	4.0
Cumulative Total.	7.5	7.5	8.2	9.3	11.5	7.4	6.8	8.0	7.1	11.0

TABLE XIV--CONTINUED

Rating Scale Letter	Case Numbers of Control Individuals									
	1	2	3	4	5	6	7	8	9	10
F (H.S. Rank and Average). . .	3.0	2.5	2.5	2.0	3.0	2.0	2.0	2.0	2.0	2.0
Cumulative Total	10.6	10.0	10.7	11.3	14.5	9.4	8.8	10.0	9.1	13.0
G (Check Lists).	1.8	1.7	2.1	2.2	2.3	1.9	2.1	2.2	2.5	1.9
Cumulative Total	12.4	11.7	12.8	13.5	16.8	11.3	10.9	12.2	11.6	14.9
H (Informal Tests)	2.3	2.7	2.7	2.6	2.4	2.4	2.4	2.7	2.4	2.4
Cumulative Total	14.7	14.4	15.5	16.1	19.2	13.7	13.3	14.9	14.0	17.3
I (Scholastic)	2.0	1.0	1.0	2.0	3.0	1.0	1.0	1.0	2.0	1.0
Cumulative Total	16.7	15.4	16.5	18.1	22.2	14.7	14.3	15.9	16.0	18.3
J (Personality Traits)	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Cumulative Total	18.7	17.4	18.5	20.1	24.2	16.7	16.3	17.9	18.0	20.3
K (Personal Appearance, Health and Hygiene).	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Cumulative Total	21.7	20.4	21.5	23.1	27.2	19.7	19.3	20.9	21.0	23.3
Mean (All Items).	2.0	1.9	2.0	2.1	2.5	1.8	1.8	1.9	1.9	2.1
Total Excluding E, H, J, K.	11.4	10.7	11.8	12.5	14.8	11.3	10.9	11.2	12.6	11.9
Mean Excluding E, H, J, K.	1.6	1.5	1.5	1.8	2.1	1.6	1.6	1.6	1.8	1.7

possessed essentially the same characteristics or abilities which the investigator desired to hold constant. The matching of each pair automatically matched the two groups. Any differences which were found in the performance of these subjects or groups at the close of the experiment could be attributed therefore to the operation of the remedial program conducted by the investigator.

The experimental qualifications of each subject were itemized on separate cards as the first step in the actual matching procedure. It has been shown in the foregoing discussion that each subject possessed the following experimental qualifications: a general prediction of potential college failure; a percentile rank of 7 or below in intelligence; a high-school average of 85 or below; a high-school rank of 86 or below; a percentile rank of 16 or below in reading comprehension; and a final rating scale score on the diagnostic instrument which gave a classification of inferior.

In order to make the experimental controls as adequate as possible, other supplementary variables were considered which previous studies¹ had shown to be of significance in conditioning scholastic success. These variables were place of residence, age, sex, and letter ratings on the Thurstone Personality Schedule. In addition, the paired subjects had to be registered for the same courses, since the comprehensive examination scores which were taken as the criterion of scholastic success had to be comparable.

The itemized cards of each subject were analyzed and tentatively matched until each experimental and control pair was equated as exactly as possible on the essential variables.

After the completion of the total matching procedure, the 23 subjects were divided into three groups, an Experimental Group A, an Experimental Group B, and a Control Group. The number of subjects in these three groups was 10, 3, and 10, respectively. Experimental Group A was given individual or clinical remedial training, a self-directed-study procedure was used with the Experimental Group B, and no remedial training was given to the Control Group. The remedial procedures used are discussed in detail in Chapter V.

¹Reeves and others, op. cit., V.

The complete results of the matching procedure are shown in Table XV. The entries in this table show that the experimental and control pairs were matched as closely and discriminatingly as possible on the essential variables. The mean differences between the experimental and control groups were 0.3 in percentile rank in intelligence, 1.5 in high-school average, 2.0 in high-school rank, 1.6 in percentile rank in reading comprehension, 0.3 in age, and 0.1 on the final rating scale score. The slight advantage that is apparent on some of the measures was given intentionally to the control group rather than to the experimental group to avoid any favorable influence on the experimental results. In the case of one measure only, that of high-school rank, the difference was in favor of the experimental group. The favorable influence of this one difference was more than compensated, however, by the differences which were favorable to the control group.

The adequacy of this final matching was of prime importance, since the reliability of the experimental results depended to a considerable extent upon the control of the variables which might condition scholastic success. The additional fact should be noted that all of the subjects used in the experiment were registered in English 102. Instruction and counsel are given in this course in "Note-taking," "Instruction in handling the commoner types of written assignments in university courses," and "Instruction in the preparation of original papers." Similarly, all of the subjects received the ordinary services of the advisory staff in the office of the Dean of students. It was necessary to control these variables in order to prevent the operation of remedial factors other than the experimental program conducted by the investigator.

Since the matching procedure involved control of the essential variables which were subject to control and which might condition scholastic success, any significant improvement attained by the experimental subjects in excess of that attained by the control subjects, at the close of the experiment, could be attributed to the experimental program.

The general experimental procedure involved the administration of the experimental variable which, in this case, was the remediation of the reading and study deficiencies of the experimental subjects. The control group was held constant throughout

TABLE XV
VARIABLES ON WHICH THE EXPERIMENTAL PAIRS WERE MATCHED

Pair	Percentile Rank in Intelligence	High School Average	High School Rank	Percentile Rank Iowa Reading Comprehension	Personality Schedule	Age	Sex	Place of Residence	Courses Registered for	Final Rating Score	
										All Items	Exclud. E, H, J, K
1-Exp. A	5	83	75	7	C	21	M*	Ho	SS, BS, E	1.6	1.4
1-Con.	5	85	82	8	C	20	M	Ho	SS, BS, E	2.0	1.6
2-Exp. A	2	82	79	1	C	19	M	Ho	SS, BS, E	2.0	1.5
2-Con.	1	83	76	1	B	19	M	D	SS, BS, E	1.9	1.5
3-Exp. A	6	85	85	7	B	17	M	Ho	SS, BS, E	2.0	1.8
3-Con.	6	84	77	6	B	18	M	Ho	SS, BS, E	2.0	1.5
4-Exp. A	6	84	85	9	C	18	M	Ho	SS, BS, E	2.2	1.9
4-Con.	6	85	68	11	C	18	M	Ho	SS, BS, E	2.1	1.8
5-Exp. A	7	84	84	6	A	17	F	Ho	SS, BS, E	2.2	2.1
5-Con.	7	83	82	7	B	18	F	Ho	SS, BS, E	2.5	2.1
6-Exp. A	4	82	84	12	B	18	F	D	SS, BS, E	1.9	1.8
6-Con.	3	84	73	13	A	18	F	D	SS, BS, E	1.8	1.6
7-Exp. A	3	83	72	10	B	18	M	D	H, BS, E	2.2	1.6
7-Con.	3	85	73	11	A	18	M	D	H, BS, E	1.8	1.6
8-Exp. A	1	81	65	6	B	18	M	Ho	H, BS, E	1.8	1.5
8-Exp. B	2	82	73	4	C	19	F	Ho	H, BS, E	1.9	1.6
8-Con.	2	84	64	6	D	18	M	Ho	H, BS, E	1.9	1.6
9-Exp. A	1	84	65	2	B	19	F	Ho	SS, BS, E	1.9	1.7
9-Exp. B	1	82	83	1	B	18	F	Ho	SS, BS, E	1.9	1.5
9-Con.	1	83	70	2	A	17	F	Ho	SS, BS, E	1.9	1.8
10-Exp. A	5	81	71	7	C	19	M	Ho	H, BS, E	2.2	1.6
10-Exp. B	7	82	52	5	A	18	M	Ho	H, BS, E	1.6	1.5
10-Con.	7	84	62	8	C	22	M	Ho	H, BS, E	2.1	1.7
Mean Exp.	3.8	82.6	74.7	6.1	B	18.3	..	..		1.9	1.6
Mean Con.	4.1	84.1	72.7	7.7	B	18.6	..	..		2.0	1.7
Mean Dif.	0.3	1.5	2.0	1.6	..	0.3	..	..		0.1	0.1

*Legend: M Male SS Social Science
F Female BS Biological Science
Ho Home E English
D Dormitory H Humanities

the experiment by the matching procedure described in this chapter and by withholding remedial training from it.

The foregoing discussion of the organization of the diagnostic and remedial experiment shows that the subjects used in the experiment were classified as potential college failures through the use of two statistical predictive methods and the classificatory portion of the diagnostic technique. It was shown further that the matching procedure used in the experiment was as discriminating as possible. The experimental results secured through the operation of the remedial program may therefore be considered statistically and experimentally reliable. The actual administration of this further diagnostic and remedial program is discussed in the next chapter.

CHAPTER V

THE EXPERIMENTAL PROGRAM OF DIAGNOSIS AND REMEDIATION

Problem of the Chapter

The major purpose of the second portion of the study is to apply the diagnostic instrument and appropriate remedial procedures to potential college failures under experimental conditions in order to measure their administrative value. Two criteria are used in this evaluation: (1) the specific improvement of the subjects in reading and study; and (2) the improvement of the subjects' scholastic performance on the comprehensive examinations above the F or failing mark which was predicted for them. The specific aim of this chapter is to describe the further diagnostic and remedial procedures used with the experimental subjects and to evaluate the results achieved in terms of the foregoing criteria.

Determination of the General Areas of Deficiency

The diagnostic materials developed in the study were designed especially to insure a diagnosis which is adequate and comprehensive. The most effective and economical use of the instrument demands a preliminary identification of the general areas of greatest weakness, followed by a more detailed study of the specific items of deficiency. This enables the remedial worker to determine at what points the process of remediation must be most intensive.

The general areas of deficiency of the subjects used in the experiment were located readily by analyzing their scores on the rating scale. These scores were presented in Table XIII. The final composite score of every subject indicated that he was inferior, as was reported in Chapter IV. A final rating of inferior does not mean, however, that the student necessarily has the same degree of inferiority on every portion of the diagnostic instrument. By projecting the scores made by a subject on each type of diagnostic material upon the proper section of the classificatory scale a separate classification may be obtained for

each type. If the score of any subject on one or more of the diagnostic materials enters the middle area of the scale relatively less emphasis may be placed in the remedial program upon the ability represented by that score than upon those abilities which have been shown definitely to be very inferior.

As reported in Table I, scores of 3.0 and 2.8 delimit the middle area of the scale on the reading tests and on the reading and study check lists, respectively, since these scores were earned by the lower middle group. By comparing the scores made by each subject on these measures with the scores made by the lower middle group, a definite classification of inferior was given to each subject who earned a score of less than 3.0 on the reading tests and less than 2.8 on the check lists. By a similar process the scores of each subject on each sub-test and check list were projected upon the proper sections of the classificatory scale and the areas of inferiority were identified. In order to illustrate the procedure more clearly the analysis of one experimental subject will be presented. For this purpose Experimental Subject A-1 is selected. He was chosen merely because he is first on the list.

The basic rating scale scores made by this subject on the diagnostic materials were as follows: Iowa Test, 1.9; Minnesota Examination, 1.0; English, 1.3; Intelligence, 1.0; Personality, 1; high-school rank, 2; high-school average, 2; reading and study check lists, 1.5; informal tests, 2.1; scholastic ratings, 1; personality traits, 2; and personal appearance, health, and hygiene, 3.

By projecting these scores upon the classificatory scale it was found that the subject was not only inferior as measured by all of these diagnostic materials, but was below the mean of the inferior group on all except high-school average, high-school rank, and the three types of personal ratings. On these measures the subject falls directly upon the mean of the inferior group. It is readily apparent, therefore, that he was deficient on all of the diagnostic materials and especially deficient in the abilities measured by the reading tests, the reading and study check lists, and the psychological examination. While his personality ratings were inferior, they were relatively less inferior than the ratings on the more academic abilities measured by the diagnostic instrument. It follows that the further study of his

specific deficiencies must be determined through the use of all of the diagnostic materials, with relatively less emphasis upon personality aspects.

The sub-areas of deficiency of subject A-1 were determined further by analyzing the scores made by him on the sub-tests and individual check lists and projecting them on the classificatory scale, in a manner similar to that used in the foregoing. It was found that the subject was deficient on all of the sub-tests of the Iowa Test with the exception of 2b, or science vocabulary. This was probably due to the fact that the subject had a science major in high school. He exhibited deficiency on all sub-tests of the Minnesota Examination, the English Test, and the American Council Psychological Examination; on all of the reading and study check lists; and on the informal individual tests with the exception of 1c (normal oral reading), 3 (pronunciation), and 12b (dot-counting), a form of general attention.

The general diagnostic status of subject A-1 may be summarized by stating that he is markedly deficient in all of the abilities measured by the diagnostic instrument with the exception of science vocabulary, general personality factors, oral reading, pronunciation and that aspect of general attention measured by dot-counting. The identification of the general areas of deficiency for all of the experimental subjects followed the same general procedure outlined in the foregoing discussion. The further procedure followed by the investigator in obtaining a more specific type of diagnosis is described in the next section of the chapter.

Determination of the Specific Items of Deficiency

The fundamental type of diagnosis used in the study has been termed an analytical diagnosis, as reported in Chapter III, because it provides an analysis of the subject's performance in the various skills involved in each area of deficiency. In order to lay the foundation for an adequate remedial program, the general areas of deficiency of each subject were subjected, therefore, to a further analysis to determine the specific items of deficiency. This was accomplished by analyzing the specific responses made by each subject on the items which compose the tests, sub-tests, and check lists of the diagnostic instrument.

The identification of the wrong responses on the test items was accomplished by checking them with the proper scoring keys. On the 19 reading and study check lists, the quality of the responses served to identify the items of deficiency. As reported earlier, these responses were rated on a five-point scale on which the degree of performance was ranked from never to always. Since each item was a good procedure or activity, as recommended by experts, and was stated in positive form, the responses of never and seldom were definite indications of deficiency. The major remedial emphasis was placed upon those check list items to which the students gave either of these responses.

The analysis of student responses to the standardized portion of the diagnostic instrument was supplemented by a similar analysis of their responses on the items of the personal and family data sheets. This type of analysis gave the investigator a fairly complete understanding of the individual student and of the manifold factors of experience, training, and environment which were contributing to or which might condition reading and study efficiency or general scholastic success. In the case of certain of the more modifiable items of home environment, part-time employment, and hygiene, remedial measures were adopted when possible or feasible.

In order to illustrate more clearly the specific diagnosis of deficiencies which may be obtained through the use of the instrument, the case of Experimental Subject A-1 is analyzed further. It was shown earlier that this subject was very deficient in all of the general factors measured by the diagnostic instrument with the exception of science vocabulary, general personality factors, oral reading, pronunciation, and that aspect of general attention which is measured by dot-counting. Further analysis of his specific responses to the Iowa Reading Test indicated that he failed to grasp the meaning of paragraphs; that his vocabulary in social science, mathematics, and English was poor; that he could not select the central idea of paragraphs or selections; that he could not analyze material and summarize it in outline form; that he could not locate information by the use of the index or by the selection of key words; and that both his reading rate and reading comprehension abilities were very poor. The deficiencies of Subject A-1 in vocabulary and reading comprehension were substantiated further by the analysis of his responses on the Minnesota

Reading Examination. His responses to the English Placement and Qualifying Examination indicated that his spelling ability was poor; that he had a tendency to use too many words or too much repetition; that his general mastery of English usage was poor; and that he could not identify the proper completion of a thought for purposes of paragraphing.

The responses made by the subject to the items on the American Council on Education Psychological Test indicated that he had very poor arithmetical ability, that his reasoning processes were slow and inaccurate, that his analysis of space relationships was poor as shown by the geometrical analogies test, and that his general vocabulary was limited.

An analysis of the responses of Subject A-1 to the items of the reading and study check lists was made to determine further the specific activities, habits, or abilities in which deficiencies were present. As shown by his responses of never and seldom to these items, he exhibited deficiency on the following: 19 items of check list number 1 on "General Physical Factors of Reading, Study, and Hygiene"; 27 items of check list number 2 on "Note-Making"; 20 items of check list number 3 on "Outlining and Summarizing"; 26 items of check list number 4 on "Making Bibliographies and Written Assignments"; 5 items on check list number 5 on "Using the Libraries"; 11 items of check list number 6 on "Attending Lecture and Discussion Sections"; 10 items of check list number 7 on "Working in the Laboratories"; 10 items of check list number 8 on "General Study Factors"; 14 items of check list number 9 on "Building a Vocabulary"; 26 items of check list number 10 on "Preparing Reading Assignments and Using Text-Books and Reference Books"; 22 items of check list number 11 on "Special Factors of Reading Rate"; 13 items of check list number 12 on "Special Factors of Reading Comprehension"; 7 items of check list number 13 on "General Reading Factors"; 3 items of check list number 14 on "Memorizing"; 6 items of check list number 15 on "Reviewing and Preparing for Examinations"; 2 items of check list 16 on "Taking Examinations"; 6 items of check list number 17 on "Developing and Maintaining Interest in Your Work"; 16 items of check list number 18 on "Thinking and Problem-Solving"; and 8 items of check list number 19 on "Factors of Habit Formation."

As indicated by the foregoing report, Subject A-1 was very deficient in 260 of the 931 items which compose the reading

and study check lists. In addition, there were 437 items on which a response of 3 or often was given, which indicates a considerable degree of deficiency. He exhibited therefore varying degrees of significant deficiency on 697 of the 931 items. On the remaining 224 items of the check list, the subject made responses of usually and always, which indicate a satisfactory degree of use or performance of the specific activity suggested by each item. From the foregoing analysis, a detailed knowledge of the specific reading and study deficiencies of the subject was obtained.

An analysis of the responses made by Subject A-1 to the items of the personal and family data sheets as well as to the questions of the Personality Schedule revealed information of considerable value in arriving at an understanding of the student's general background. It was found that the subject's father is a salesman who makes approximately \$2,000.00 a year, and who is divorced from the mother. Since the granting of the divorce four years ago, the subject has lived with the mother and has worked in an office for three hours a day to contribute to the family support. Two younger siblings are in high-school. The mother completed high-school and the father completed college. Both parents are very nervous and irritable, and controlled the subject during childhood mainly by corporal punishment. The main recreation of the entire family is the "Movies." Until the time of the divorce the entire family traveled in the United States during summer vacations.

Subject A-1 is of English descent, male and 21 years of age. He was born and educated in a city of about 20,000 population. Having never attended nursery school or kindergarten, he was unable to read when he entered elementary school. He was never encouraged to read as a child, read literature analytically in his high-school English courses, and had reading difficulties in both elementary school and high-school, as indicated by the fact that he experienced difficulty in covering and understanding his reading assignments. The subject felt that he read too slowly and did not read thoroughly, in general.

Approximately three hours are spent in study by Subject A-1 on week-days and Saturdays, and one hour on Sundays. He studies best in the early morning and usually retires early. The subject spends approximately 20 minutes per day in commuting to school, about one hour per day in listening to the radio, and 30

minutes in reading the daily newspapers. He enjoys reading but prefers fiction. The subject lived at home throughout his school career and still lives with his mother and siblings in a small and rather crowded apartment.

Subject A-1 pursued a biological science major in high school and is continuing it at present, having chosen that division of specialization for his college and graduate work. He considers fame as the important measure of success and has no definite philosophy of life. His college entrance age was 21 years, owing to the fact that he worked in an office for two years after graduating from high-school.

His personality is rather unstable, vacillating, and subject to emotional outbursts. He had had religious and sex conflicts which disturbed him considerably. He day-dreams, broods, and worries much; in addition he is temperamental, imaginative, and excessively afraid of disease. The subject has had considerable sickness, including influenza and a major operation for appendicitis.

Consideration of the facts presented in the foregoing paragraphs indicates that Subject A-1 is not only deficient in general reading and study abilities, but has a considerable degree of maladjustment with respect to other phases of his life and environment. For example, the crowded condition of his home is unfavorable to efficient study. It was recommended therefore that he study more in the university libraries. The student's statements to the effect that he had reading difficulties in high-school and that he felt that he read slowly and not thoroughly, provided a point of departure in gaining his cooperation. It was recommended further that he study more on Sundays, and devote less time to listening to the radio. Similarly, a consideration of the other facts revealed by this supplementary analysis gave the investigator a general understanding of the subject, and assisted him in becoming "en rapport" with the student much more effectively than could have been accomplished without a knowledge of his general life and personality.

The diagnostic procedure followed in the case of Subject A-1 is typical of that followed with the other experimental subjects. Through general and specific analyses, a comprehensive basis was laid for the subsequent program of remediation. Before discussing the remedial procedures adopted, however, one further

diagnostic step will be described which is not an essential part of the instrument developed in the study but was used in the experimental portion of the investigation because of its availability and objectivity.

Use of the Eye-Movement Camera

One of the basic laboratory devices used widely during the last twenty years in the investigation of reading is the eye-movement camera. This device enables an investigator to obtain an objective record, in the form of a photographic film, of the three fundamental types of eye-movements which occur during the reading process. These measures are: (1) average number of fixations per line; (2) average duration of fixation pauses; and (3) average number of regressive movements per line.

In the experimental portion of the study the camera was used to obtain an objective record of the subjects' eye-movements at the beginning and close of the experimental training period. In addition to supplying evidence of the amount of improvement, the records were very valuable in demonstrating to the subjects, objectively, their actual reading deficiency in these three fundamental aspects of the reading process. They thus served as a means of stimulating or motivating the program of remediation.

Eye-movement photographs were secured for each of the 23 subjects used in the experiment. In addition, the records of ten superior readers and ten average readers were obtained in order to establish a simple basis of comparison by means of which a general interpretation could be made of the performance of the experimental subjects. The superior readers comprised the 98 to 100 percentiles on the Iowa and Minnesota Reading Tests, while the average readers comprised the 49, 50, and 51 percentiles on the same tests. The reading selection used while photographing the eye-movements of the subjects was taken from the Social Science Syllabus.¹ It was a typical yet simple sample of college reading materials. The selection follows:

¹W. Bagehot, "Physics and Politics," Syllabus and Selected Readings, Introductory General Course in the Social Sciences, XXVI, 412. Second Preliminary Edition. Chicago: University of Chicago Press, 1932.

"The progress of man requires the co-operation of men for its development. That which any one man or any one family could invent for themselves is obviously exceedingly limited. And even if this were not true, isolated progress could never be traced. The rudest sort of co-operative society, the lowest tribe and the feeblest government, is so much stronger than isolated man that isolated man might very easily have ceased to exist in any state whatsoever.

The subjects were instructed to read this selection at their ordinary reading speed and to secure a general understanding of the content or meaning of the passage as they read. The general construction and operation of the eye-movement camera has been described at length by G. T. Buswell.¹

The tabulated results of the records obtained for these subjects on the three fundamental types of eye-movements are shown in Table XVI. The data show that the scores of the individual subjects used in the experiment and the mean scores of the experimental and control groups were inferior to the mean scores of the average and superior readers. They show further that the subjects had the same general type and degree of deficiency. Each subject had an average number of fixations and regressive movements per line which was greater than the mean of the superior and average readers and an average duration of fixation pauses which was longer than that of the same readers. It is obvious, therefore, that the measures derived from the use of the eye-movement records supplemented to advantage the general and specific diagnosis made through the use of the diagnostic materials described in Chapter III.

Upon the basis of the foregoing diagnosis of the deficiencies of the experimental subjects a comprehensive program of remediation was organized for the improvement of reading and study abilities.

The Basic Remedial Materials

A diagnosis is of little value unless remedial procedures are adopted to overcome or correct the deficiencies identified by

¹G. T. Buswell, A Laboratory Study of the Reading of Modern Foreign Languages, p. 7. New York: The Macmillan Co., 1927.

TABLE XVI
SCORES OBTAINED FROM PHOTOGRAPHIC RECORDS OF THE EYE-
MOVEMENTS OF THE EXPERIMENTAL AND CONTROL SUBJECTS

Subject	Average Number of Fixations Per Line	Average Dura- tion of Fixa- tion Pauses	Average Number of Regressive Movements Per Line
Mean score of ten good readers	4.5	6.1	0.04
Mean score of ten average readers	5.3	6.0	0.2
Experimental subjects.			
A 1	7.1	6.4	0.8
2	7.1	6.8	0.7
3	6.8	7.1	0.6
4	7.2	6.7	0.9
5	7.3	6.6	0.5
6	6.4	7.4	0.9
7	6.8	7.0	0.5
8	6.8	6.9	0.7
9	6.9	6.8	0.8
10	6.8	6.7	0.6
B 1	6.6	6.4	0.8
2	6.9	6.9	0.9
3	6.8	7.0	0.6
Control 1	6.6	6.7	0.6
2	7.1	6.5	0.8
3	6.8	6.9	0.7
4	6.9	7.0	0.7
5	6.7	6.9	0.5
6	6.8	6.8	0.5
7	7.0	6.7	0.6
8	6.9	6.4	0.9
9	7.0	6.7	0.7
10	7.1	6.6	0.8
Mean Exper. A. . . .	6.9	6.8	0.7
Mean Exper. B. . . .	6.8	6.8	0.8
Mean Control	6.9	6.7	0.7

the use of the diagnostic measures. Since the chief purpose of the experimental portion of the study was to apply the diagnostic instrument to potential college failures under experimental conditions, the remedial procedures used to overcome the deficiencies identified during the diagnosis are of major importance. The two types of remedial materials constructed and used in the study and the general remedial program adopted will be described in the succeeding sections of the chapter.

Use of the comprehensive reading and study check lists as remedial guides to the best procedures.--The fact was pointed out earlier in the report that the reading and study check lists developed in the study were composed of 931 specific items or activities which were recommended by experts as good procedures in reading, study, and hygiene. It was shown further that the analysis of the student responses to these items enabled the investigator to identify the items in which deficiency was present. Since each activity in the check list was a good procedure and was stated in positive form, the lists may be used further as comprehensive remedial guides to the best procedures. For example, if the response of a student to check list 2B, item 2, which is "Make notes in ink, preferably black," was either never or seldom, the deficiency is apparent and the remedial procedure is suggested by the positive statement. It is apparent, therefore, that if the positive statements which compose the check list are followed out by the subjects they will be making use of the best and most efficient reading and study procedures. This use of the check lists furnished the major basis for the remedial program adopted in this investigation. Before entering upon the actual presentation of this program of remediation, it is necessary to describe a supplementary type of remedial material which was constructed especially for the study but which has not been described up to this point.

Use of the vocabulary lists in the four general survey courses.--Previous investigators¹ have shown that there is a definite positive relationship between vocabulary and reading

¹Arthur E. Traxler, "The Measurement and Improvement of Silent Reading at the Junior-High-School Level." Unpublished Doctor's Dissertation, Department of Education, University of Chicago, 1932.

achievement. Many of the investigators,¹ however, when faced with the need for improving reading vocabulary, have used published word lists, such as the Thorndike² "Word Book," as the basis for the remedial practice exercises. Since most of the published word lists were derived from an analysis of the word usage of selected cross-sections of the general population they represent common usage³ rather than the technical usage involved in the academic readings with which college students must deal.

It was shown earlier that all of the subjects used in the experiment were deficient in vocabulary, as measured by the Iowa and Minnesota Reading Tests. The investigator therefore constructed four vocabulary lists which are closely integrated with the academic demands made by the "New Plan" at the University of Chicago. The four "New Plan" General Survey Courses furnished a splendid source of materials for the construction of vocabulary lists which might satisfy the more technical and specialized needs of college students. These courses are catalogued as the General Introductory Courses in the Humanities, the Social Sciences, the Biological Sciences, and the Physical Sciences. They cut across departmental lines and attempt to give a complete and integrated over-view of all of the major fields of human knowledge.

A basic vocabulary list was constructed in each of these courses after tabulating the technical and difficult words which were identified during a critical survey and analysis of two sources: (1) the general syllabus and selected readings for each survey course; and (2) the comprehensive examinations which had been given in these courses over a period of seven quarters, June, 1932 to September, 1933, by the Board of Examinations.

From the analysis of these basic materials approximately 4,500 technical and difficult words were selected by the investigator. These were submitted to graduate students in each of the

¹Henry T. Tyler, "Remedial Training in the Junior College," The Junior College Journal, VI (October, 1933), 28-31.

²Edward L. Thorndike, A Teachers Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People. New York: Bureau of Publications, Teachers College, Columbia University, 1931.

³Ibid., pp. 165-82.

four divisions to determine the accuracy of their spelling and usage. They were submitted in turn to 20 freshmen, selected at random, who checked the words which they felt caused some difficulty in comprehension. The 3,700 words which the graduate students approved and which the majority of the freshmen indicated as having caused some difficulty in comprehension were finally incorporated into four lists, one for each course. Over-lapping between the lists was avoided, as far as possible.

These lists served two major purposes: (1) to furnish a guide to the basic vocabulary necessary for efficient reading and study in the four general courses; and (2) to focus the student's attention on the basic concepts necessary for successful performance on the final comprehensive examinations.

The procedure followed in using the lists as remedial instruments was identical for all experimental subjects. No attempt was made to actually teach the words to the subjects. A self-directed-study method was adopted which involved the following steps: (1) to make a preliminary survey of the lists and check all of the words which were thoroughly familiar; (2) to check, at the end of each week, the words that had appeared in the readings and lectures of the week and to focus attention on those terms, making certain that their knowledge of the words was adequate and fairly complete; (3) to use the lists as measures of progress in the courses by noting the relative number of words or concepts which had been mastered at various stages of the courses; and (4) to use the lists in securing final measures of their knowledge of the content of the courses before taking the final comprehensive examinations.

In serving their purpose as instruments for training in vocabulary, the lists provided a valuable supplementary aid to the general program of remediation in reading and study. The major remedial instrument was the check lists, however, rather than the vocabulary lists. On the basis of these two types of remedial materials a comprehensive remedial program was organized and conducted during the academic year 1933-34, beginning October 23 and ending June 2.

The General Remedial Program

The Control Group was equated with the Experimental Group on all of the essential variables to hold constant the operation

of factors other than the experimental program of remediation. It was necessary, therefore, to rule out this final variable by giving remedial training to the experimental subjects and withholding it from the control subjects. Any subsequent improvement which might be noted in the experimental subjects in excess of that achieved by the control subjects could therefore be attributed to the operation of the remedial program.

The "New Plan" at this institution involves freedom from compulsory class attendance. It was deemed desirable therefore to obtain some measure of the degree to which students of the type used in the study are capable of efficient and economical self-directed-study activity. After the administration of the diagnostic materials and subsequent analysis of the items of deficiency, as described earlier, the students in Experimental Group B were used in obtaining a measure of this type of remedial procedure. Each subject was given a complete set of the reading and study check lists on which the items of deficiency were properly identified and the use of the lists as remedial devices and as guides to the best procedures in reading and study was explained. They were provided, in addition, with the vocabulary lists and instructions for their use, as suggested on the preceding page. These subjects were dismissed after the initial conference and were not summoned again until the administration of the final testing program.

The major portion of the remedial program, however, was conducted with the Experimental Group A. Intensive individual conferences were held with these 10 subjects for a period of 29 weeks. The reading and study check lists furnished the basis for the conferences. In each conference the specific items of deficiency which the subject exhibited on the lists were discussed and the proper remedial suggestions were indicated, as suggested by the statements of good procedure made in the lists. A sequential and formal report of these conferences is presented here for two reasons: (1) to set up a tentative program of conferences which might be followed by future investigators or remedial workers; (2) the remedial procedure used with each subject was formalized to the extent that the same check list was used throughout the conferences of any particular week. The remedial procedures would vary, however, with the items of deficiency which each subject exhibited on that particular list. Since no two subjects

exhibited the same combinations of deficiencies the remedial procedures were provided on an individual rather than a group basis.

The complete program of remedial conferences conducted with Experimental Group A is summarized in the following list:

Week of:

- November 13-18 Fourth conference. (1 hour) (The first three conferences were concerned with organizing the experiment and administering the diagnostic instrument.)
The vocabulary lists were given to the subjects with the proper instructions for their use. Check list number 9, "Building a Vocabulary," was given to each subject. The specific items of deficiency were indicated and the proper remedial procedures were suggested. The responses of the students to the items of the data sheets were discussed with them and remedial suggestions were made where possible or feasible, especially in regard to making home conditions more desirable for study purposes.
- November 20-25 Fifth conference. (1 1/2 hours) (At the students' homes or rooms.)
The subjects were given check list number 1, "General Physical Factors of Reading, Study, and Hygiene." The specific items of deficiency were indicated and the proper remedial suggestions were made. The subjects were advised further concerning the general betterment of home conditions, reading and study conditions, hygiene, and diet.
- November 27-December 2 Sixth conference. (1 hour)
The subjects were given check list number 2, "Making Notes." The items of deficiency were indicated, remedial suggestions were made, and the best note-making procedures were presented. Suggestions on note-making were confined to objective activities or procedures in the main, rather than to factors of correct English usage, since the advisers in English 102 emphasized the more qualitative aspects of note-making. The students' usage of the vocabulary lists was checked informally.
- December 4-9 Seventh conference. (1 hour)
The subjects were given check list number 5, "Using the Libraries." The subjects were taken to the libraries, the use of general and special library aids was demonstrated, and general library procedures were indicated.
- December 11-16 Eighth conference. (1 hour)
The subjects were given check list number 8, "General Study Factors." The items of deficiency were indicated and remedial suggestions were made. The usage of the vocabulary lists and note-making procedures were checked informally.

- December 18-22 Ninth conference. (1 hour)
The subjects were given check list number 6, "Attending Lecture and Discussion Sections." The items of deficiency were indicated and specific remedial suggestions were made. The changes made by the students in general study factors, procedures, and conditions were noted informally.
- January 2-6 Tenth conference. (1 hour)
The subjects were given check lists number 7 and 18, "Working in the Laboratories," and "Thinking and Problem-Solving." The items of deficiency were indicated and specific remedial suggestions were made. An informal check was made on the subjects' note making, library usage, and reading and study hygiene.
- January 8-13 Eleventh conference. (1 hour)
The subjects were given check lists number 3 and 4, "Outlining and Summarizing," and "Making Bibliographies and Written Assignments." The specific items of deficiency were indicated and remedial suggestions were made. Objective procedures or activities were emphasized, rather than the more qualitative aspects, since the advisers in English 102 emphasize the factors of English usage, documentation, etc. An informal check was made of the students' usage of the remedial suggestions in check list number 6.
- January 15-20 Twelfth conference. (1 hour)
The subjects were given check lists number 14, 17, and 19, on "Memorizing," "Developing and Maintaining Interest in Your Work," and "Factors of Habit Formation," respectively. The specific items of deficiency were indicated and remedial suggestions were made. An informal check was made on the progress of the students in note-making, written papers, outlines and summaries, and vocabulary.
- January 22-27 Thirteenth conference. (1 hour)
The subjects were given check list number 10, "Preparing Reading Assignments and Using Text-Books and Reference Books." The items of deficiency were indicated and specific remedial suggestions were made. Drill was given on the use of special parts of books, such as the index, table of contents, tables, diagrams, figures, foot-notes, maps, etc.
- January 29-February 3 Fourteenth conference. (1 hour)
The subjects were given check list number 13, "General Reading Factors." The specific items of deficiency were indicated and remedial suggestions were made. An informal check was made of the students' use of the remedial suggestions contained in check list number 10.
- February 5-10 Fifteenth conference. (1 hour)
The subjects were given check list number 12, "Special Factors of Reading Comprehension." The items of deficiency were indicated and remedial suggestions were

made. Each subject was given the results of the placement tests in reading to stress the need for remedial training and to serve as a factor of motivation.

- February Sixteenth conference. (1 hour)
12-17 The subjects were given check list number 11, "Special Factors of Reading Rate." The specific items of deficiency were indicated and remedial suggestions were made. Each subject was given the results of the placement tests in reading rate and the results of the eye-movement photographs to emphasize the need for remedial training in reading rate.
- February Seventeenth conference. (1 hour)
19-24 The subjects were given specific drill or practice on materials in the college syllabi and selected readings designed to increase reading comprehension, such as picking out topic sentences, key phrases and sentences, answering questions about the content of the material read, and summarizing the material read. The subjects were introduced to the tachistoscopic or short-exposure method of increasing span of recognition or perception.
- February Eighteenth conference. (1 hour)
26-March 3 The subjects were given specific drill or practice designed to increase reading rate, such as skimming 20 pages of material and noting only the paragraph headings, key statements, etc. Short exposure drill was given on sentences to increase the span of recognition. Subjects were forced to read materials at the rate of 400 to 600 words per minute. The subjects read an easy selection from the Social Science Syllabus over and over, increasing the speed at each successive reading.
- March Nineteenth conference. (1 hour)
5-10 Continued drill was given to the subjects in reading rate and comprehension, such as checking the number of pages of material read in 10 minutes, skimming materials and writing down all that could be recalled, summarizing materials read, etc. The subjects were asked to outline reading materials for the lectures of the following week. Additional short-exposure and rapid reading drills were given.
- March Twentieth conference. (1 hour)
12-17 Drill on reading rate and comprehension was continued, such as getting a question in mind and skimming materials to find an answer, following test directions, and analyzing analogy type statements. Drill was given on reading materials with wide spacing between words and marks placed above key words to stimulate rhythical and uniform eye-movements along the lines.
- March Twenty-first conference. (1 hour)
19-24 Drill was continued on reading rate and comprehension, such as following test directions, skimming materials to summarize them or to give topic sentences, reading

certain materials to outline them, etc. The outlines of the lectures of the preceding week were checked for form. Forced rapid-reading and short-exposure drills were given.

- April 2-7 Twenty-second conference. (1 hour)
An informal check was made of notes, outlines, written papers, and other student materials for form, adequacy, and correctness of procedure. A check was made also of the students' use of the vocabulary lists, of libraries, and of their reading and study conditions and hygiene.
- April 9-14 Twenty-third conference. (1 hour)
This conference was used as a period of general discussion with the subjects concerning their problems and progress. Each subject was asked to bring up any specific questions or suggestions that he might have, or any criticism of the training provided. An informal evaluation of progress was made cooperatively by the students and the investigator.
- April 16-21 Twenty-fourth conference. (1 hour)
The subjects were given check list number 15, "Reviewing and Preparing for Examinations." The items of deficiency were indicated and the proper remedial suggestions were made. An informal check was made of the subjects' use of the memorizing procedures suggested in check list number 14.
- April 23-28 Twenty-fifth conference. (1 hour)
A further check was made of the students' memorizing procedures and problem-solving procedures in the light of the suggestions contained in the check lists. The best methods of preparing for and answering the different types of examination questions were indicated.
- April 30 -May 5 Twenty-sixth conference. (1 hour)
An informal check was made of the subjects' knowledge of the words and concepts in the vocabulary lists to furnish indices of readiness for the comprehensive examinations and to emphasize the need for concentrated review.
- May 7-12 Twenty-seventh conference. (1 hour)
The subjects were given check list number 16, "Taking Examinations." The specific items of deficiency were indicated and remedial suggestions were made. Specific drill was given on previous examinations, including the rapid interpretation of test directions and the actual answering of questions. The drill was based on the "Manual of Test-Making," and on complete examinations in the four fields covered in the general survey courses.
- May 14-19 Twenty-eighth conference. (1 hour)
Continued drill was given on the best procedures in taking examinations, on test directions, and on answering actual examination questions.

The foregoing program of conferences shows the general nature of the remedial procedures that were used with the Experimental Group A. It is apparent from a consideration of these reports of the conferences that the program of remediation was extensive in scope and intensive in nature. Approximately 26 1/2 hours of help and guidance were given to each subject in this group. Only one hour was given to the subjects in Experimental Group B, and none to the Control Group.

The subjects responded satisfactorily to the program when they were shown the actual need for the work and the possible scholastic penalty for non-attendance at the conferences. They were assured that the remedial training would, in all probability, be of considerable personal and scholastic value. Some absences from conferences were inevitable, but they were relatively rare and the work of the conferences that were missed was recapitulated in later conferences.

The actual program of remediation ended May 19. During the period beginning May 21 and ending June 2 the final testing program was administered to both the experimental and control groups to determine the gains made during the experimental period.

The Final Testing Program

In order to evaluate the improvement of the subjects in reading and study that resulted from the experimental program, it was necessary to conduct a final testing program which included the same measures administered in the initial testing program. The complete battery of diagnostic materials was not used, owing to the fact that few of the tests had duplicate forms. Only those test materials were included in which specific improvement might be expected. Those used in the final testing program were: the Iowa Silent Reading Test, Form B; the Minnesota Reading Examination, Form A; the 19 reading and study check lists; the informal tests, 1, 5, 6, 7, 8, 9, 10; and the eye-movement camera.

The experimental and control subjects were summoned individually to the investigator's office, where the tests and lists were administered. The results of the final testing were then tabulated and comparisons were made between the performance of each group on the initial and final tests to determine the gains made during the experiment. In addition, the marks earned on the comprehensive examinations were obtained from the Board of Exami-

nations, and a study was made of the scholastic success achieved by the three groups. The final testing program supplied data therefore by means of which the value of the experimental program could be measured in terms of the improvement of the subjects in reading and study, as well as in scholastic achievement.

Summary of the Results of the Experiment

The diagnostic instrument and related remedial procedures developed and used in the study would have little value or utility, except as academic curiosia, if their use did not result in the actual improvement of reading and study abilities and of scholastic success. The results of the diagnostic and remedial experiment are evaluated, therefore, at this point.

The actual raw scores earned on the initial and final tests by the Experimental Group A, Experimental Group B, and Control Group are presented in Table XVII. They include the basic data upon which computations and comparisons are made later in more significant terms. The entries in this table show that the raw scores made by the experimental groups on the final tests were considerably higher than the scores earned by the same groups on the initial tests. It is shown further that the gains made by the Control Group were relatively small. The actual significance of these gains can not be evaluated readily, however, through a consideration of the raw scores alone. For example, a gain of 4 points on a scale which consisted of 10 points is very different from a gain of 4 points on a scale which consisted of 100 points. In order to reduce the gains to more comparable and understandable terms, the gains in raw score were converted into percentage gains. These data are shown in Table XVIII.

The entries in this table show that the percentage gains for each group varied considerably on the different test materials. The range in percentage gains for the three groups was as follows: Experimental A, 216.8 to 10.3; Experimental B, 262.5 to 5.3; and the Control, 40.2 to -14.3. The gain of 262.5 per cent made by Experimental Group B on one measure, the Minnesota Examination, is due not so much to the size of the gain as to the fact that the initial score of this group was considerably lower than that of the other two groups. The actual raw score gain made on this test by Experimental Group A was 38.8, while that of Experimental Group B was only 21.0. These inadequacies of the initial matching

TABLE XVII
COMPARATIVE RAW SCORES OBTAINED ON THE INITIAL AND FINAL
TESTS BY THE EXPERIMENTAL AND CONTROL GROUPS

Measure Used	Mean Scores on the Initial Test			Mean Scores on the Final Test		
	Exp. A	Exp. B	Control	Exp. A	Exp. B	Control
Iowa Comprehension.	111.5	103.0	112.4	164.3	142.8	119.2
Iowa Rate	23.3	22.3	24.9	41.5	32.1	26.7
Minn. Vocabulary. .	17.9	8.0	18.4	56.7	27.0	25.8
Minn. Paragraph . .	8.9	4.0	9.2	16.2	10.9	10.7
Reading and Study Check Lists: . . .	2976.7	2996.0	2976.3	4203.9	3732.7	3077.4
Informal Tests						
1a.	92.4	92.3	91.6	66.4	78.6	88.3
1b.	56.8	55.0	54.4	45.2	48.5	53.3
1c.	50.9	52.6	53.1	43.0	49.8	50.5
5	16.7	17.3	16.2	23.5	19.7	16.4
6	17.6	17.0	17.5	22.3	20.3	17.9
7	17.1	17.3	16.6	22.0	19.7	16.9
8	12.1	13.7	12.4	16.9	15.4	12.7
9	6.4	5.7	6.7	10.4	8.3	7.1
10	18.6	17.3	16.5	24.3	20.5	17.1
Eye-movements:						
Average number of fixations per line	6.9	6.8	6.9	5.5	6.1	6.8
Average duration of fixations per line	6.8	6.8	6.7	6.1	6.4	6.7
Average number of regressive move- ments.	0.7	0.8	0.7	0.2	0.5	0.8

on the various measures are smoothed out somewhat when the means gains on all of the test materials are computed. The mean percentage gains of the three groups on all measures were: Experimental A, 54.3; Experimental B, 44.2; and Control, 5.8. The percentage gains on all of the test materials for Experimental Groups A and B were therefore 9.4 and 7.6 times, respectively, the gain made by the Control Group on the same tests.

Since the Control Group was matched adequately on all of the essential variables with the Experimental Groups, as shown in Chapter IV, the gains made throughout the training period by the Control Group may be considered as a measure of the chance variation or maturation which occurred during that period. When eval-

TABLE XVIII
RAW SCORE AND PERCENTAGE GAINS MADE BY THE EXPERIMENTAL
AND CONTROL GROUPS DURING THE DIAGNOSTIC AND
REMEDIAL EXPERIMENT

Measure Used	Gain in Raw Score and Per Cent					
	Exper. A		Exper. B		Control	
	Raw	Per Cent	Raw	Per Cent	Raw	Per Cent
Iowa Comprehension. .	52.8	47.4	39.8	38.7	6.8	6.0
Iowa Rate	18.2	78.1	9.8	44.0	1.8	7.2
Minn. Vocabulary. . .	38.8	216.8	21.0	262.5	7.4	40.2
Minn. Paragraph . . .	12.5	144.5	6.9	172.5	1.5	16.3
Reading and Study Check Lists	1227.2	41.2	736.7	24.3	101.1	3.4
Informal Tests:						
1a.	26.0	28.1	13.7	14.9	3.3	3.6
1b.	11.6	20.4	6.4	11.6	1.1	2.0
1c.	7.9	13.6	2.8	5.3	2.6	4.9
5	6.8	40.7	2.4	13.9	0.2	1.2
6	5.1	29.0	3.3	19.4	0.4	2.3
7	4.9	28.7	2.4	13.9	0.3	1.8
8	4.8	39.7	1.7	12.4	0.3	2.4
9	4.0	62.5	2.6	45.6	0.4	6.0
10	5.7	30.6	3.2	18.5	0.6	3.6
Eye-movements:						
Average number of fixations per line	1.4	20.3	0.7	10.3	0.1	1.4
Average duration of fixations per line	0.7	10.3	0.4	5.9	0.0	0.0
Average number of regressive move- ments per line . .	0.5	71.4	0.3	37.5	-0.1	-14.3
Mean Gain. . . .		54.3		44.2		5.8

uated on this basis, therefore, the gains made by the Experimental Groups were significantly greater than that made by the Control Group.

There remains, however, a further difficulty in the interpretation of percentage gains, similar to that which occurs in considering raw scores. To take an extreme example, a raw score gain of 1 point over an initial score of 1 gives a final score of 2, which when converted into a percentage gain, is 100. It is readily apparent that such a gain is not comparable with a gain

of 100 per cent achieved over a basal or initial score of 25. In the latter case the gain in raw score should be 25 points to equal a percentage gain of 100.

In order to overcome this difficulty in interpretation, which is inherent in the percentage method, use was made of the comprehensive rating scale. By projecting the initial and final test scores of the subjects upon this scale and rating them on the five-point basis, the disproportionate percentage gains were reduced to more significant terms. The results of this treatment of the raw scores are shown in Table XIX.

TABLE XIX
GAINS MADE BY THE EXPERIMENTAL AND CONTROL GROUPS
ON THE COMPREHENSIVE FIVE-POINT RATING SCALE

Measure Used	Initial Rating			Final Rating			Gain		
	Exp. A	Exp. B	Con.	Exp. A	Exp. B	Con.	Exp. A	Exp. B	Con.
Iowa Comprehension.	1.8	1.7	1.8	3.8	2.7	1.9	2.0	1.0	0.1
Iowa Rate	2.1	1.9	2.3	4.9	3.7	2.5	2.8	1.8	.2
Minn. Vocabulary. .	1.6	0.8	1.6	4.3	2.3	2.2	2.7	1.4	.6
Minn. Paragraph . .	1.4	0.7	1.5	2.8	1.7	1.6	1.4	1.0	.1
Reading and Study Check Lists . . .	2.5	2.6	2.5	6.4*	5.4	2.8	3.9	2.8	.3
Informal Tests:									
1a.	2.8	2.8	2.9	5.3*	4.3	3.2	2.5	1.5	.3
1b.	3.5	3.7	3.8	4.9	4.5	4.0	1.4	0.8	.2
1c.	3.5	3.1	3.0	4.6	3.8	3.6	0.9	0.7	.6
5	2.3	2.5	2.3	4.4	2.9	2.3	2.1	0.4	.0
6	2.5	2.3	2.5	4.6	3.1	2.6	2.1	0.8	.1
7	2.6	2.7	2.5	3.7	3.2	2.6	1.1	0.5	.1
8	2.5	2.9	2.6	4.3	3.7	2.7	1.8	0.8	.1
9	2.2	1.8	2.3	3.7	3.3	2.5	2.5	1.5	.2
10	2.3	2.1	1.9	4.1	2.7	2.0	1.8	0.6	0.1
Mean.	2.4	2.3	2.4	4.5	3.4	2.6	2.1	1.1	0.2

*It was necessary to extend the original five-point scale in order to compute this gain. This was done by interpolating the difference between the upper end of the five-point scale and the total possible score on the test.

The entries in the table show that the mean gains made on the five-point scale by the three groups were as follows: Experimental Group A made a mean gain which is slightly less than half of the entire area of the five-point scale. As reported earlier,

the gains made by the Control Group may be considered as a measure of the chance variations during the training period, and the significance of the gains made by the Experimental Groups may be shown by computing the degree to which they exceed this chance variation. In these terms the gain made by Experimental Group A was 10.5 times its chance variation and the gain made by the Experimental Group B was 5.5 times its chance variation. By a similar computation, the relative value of the individual or clinical remedial program over the self-directed-study procedure was estimated readily. The difference between the final rating scale gains of these two groups was 1.0 in favor of Experimental Group A. This difference is 5.0 times its chance variation.

The greatest gains made by the Experimental Groups were on the reading and study check lists. These gains were 3.9 and 2.8 for Groups A and B, respectively. Such results were expected, however, since the remedial program was based primarily upon the check lists. The gain made by the Control Group on this same measure was 0.3. It is apparent therefore that the reading and study habits of inferior students do not improve automatically or without specific training to any great extent during their college years. It follows that the value of the experimental program is of considerable significance as a means of improving the personal efficiency of inferior students in reading and study. Whether or not this increased personal efficiency was reflected in improved scholastic performance is the final major consideration.

The marks earned by the subjects on the comprehensive examinations at the close of the diagnostic and remedial experiment were used to measure the value of the experimental program as a means of student retention. As reported in Chapter IV, a statistical prediction of F or failure had been made for all of the subjects used in the experiment. In addition, the marks actually earned by students of the two previous "New Plan" classes, 1931-32 and 1932-33, who fell into the same predictive categories as the subjects used in the experiment, were obtained as a basis for comparison. These data were reported originally in Table XII and are reproduced for comparative purposes in Table XX, which shows, in addition, the actual marks earned on the comprehensive examinations by the Experimental and Control Groups.

TABLE XX

SCHOLASTIC RESULTS ATTAINED BY THE EXPERIMENTAL AND CONTROL GROUPS ON THE COMPREHENSIVE EXAMINATIONS AT THE END OF THE DIAGNOSTIC AND REMEDIAL EXPERIMENT AND BY SIMILARLY SELECTED GROUPS IN THE TWO PREVIOUS "NEW PLAN" CLASSES

Mark	Percentage of Marks Earned by Two Similar Groups, 1931-32, 1932-33	Exper. A		Exper. B		Control	
		No.	Per Cent	No.	Per Cent	No.	Per Cent
* A . . .		1	3.3	...		...	
B . . .		5	16.7	1	11.1	...	
C . . .	5.4	7	23.4	2	22.2	1	3.3
D . . .	16.4	13	43.3	2	22.2	7	23.4
F . . .	78.2	4	13.3	4	44.5	22	73.3
Total	100.0	30	100.0	9	100.0	30	100.0

The data in this table show that, despite the general prediction of F or failure which had been made for these subjects, the following percentages of F marks were actually earned: Experimental A, 13.3; Experimental B, 44.5; and the Control, 73.3. The percentage of F's, or failing marks, earned by similar groups in the two previous classes, 1931-32 and 1932-33, was 78.2. The actual gain or improvement in scholastic performance made by the three groups may be represented, therefore, by the difference between the percentage of marks earned by previous groups selected on the same predictive criteria and those earned by the three groups used in the experiment. These percentage differences, which represent a decrease in the number of F marks actually earned, were 64.9 for the Experimental Group A, 33.7 for the Experimental Group B, and 4.9 for the Control Group. Using the percentage gain made by the Control Group as a measure of the chance variation, it was found that the Experimental Groups A and B made scholastic gains which were 13.2 and 6.9 times the chance variation. While the two previous groups, which were selected on identically the same predictive criteria, earned no A or B marks, Experimental Group A made 3.3 per cent A's and 16.7 per cent B's. Experimental Group A also made scholastic gains which were 31.2 per cent better than those made by Experimental Group B. This difference is 6.4 times its chance variation or the gain made by

the Control Group. Owing to the small number of subjects in Experimental Group B, however, no general conclusions will be drawn concerning the relative superiority of the clinical program of remediation used with Experimental Group A. Such general conclusions would be statistically unreliable.

It is apparent from the data presented in Table XX that the experimental program conducted by the investigator was of significant value in promoting student retention. Since the materials developed in the study were the basis for this program of diagnosis and remediation, their administrative value has been clearly demonstrated.

In consideration of the facts presented in this chapter which demonstrate the value of the experimental program as a means of improving personal efficiency in reading and study as well as scholastic performance, the further use of the instrument is recommended in the concluding chapter.

CHAPTER VI

FINDINGS AND INTERPRETATIONS OF THE STUDY

This study has been concerned with the measurement and improvement of reading and study abilities at the freshman level in college. Two fundamental problems have been investigated: (1) the development and partial standardization of a comprehensive diagnostic instrument in reading and study, which makes possible the early differentiation between potentially superior and inferior students, for purposes of subsequent remediation of the latter; and (2) the application of the diagnostic instrument and related remedial procedures under experimental conditions to measure their administrative value. The purpose of this chapter is to summarize the findings of the study, to present the general conclusions that can be drawn from them, to indicate the contributions and limitations of the study, and to give suggestions for the use and further study of the technique.

The Development and Use of the Comprehensive Diagnostic Instrument

A critical analysis of previous related investigations revealed the following facts: that a considerable portion of our college population is deficient in reading and study abilities; that a significant positive relationship exists between scholastic success and reading and study abilities; that college administrators are attempting to adjust the individual student to the institution, using diagnostic and remedial procedures in reading and study as one of the major phases of adjustment; and that there is no adequate diagnostic instrument in reading and study available at the college level. Consideration of these facts led the investigator to undertake the development and partial standardization of an adequate and comprehensive diagnostic instrument which would be of actual administrative value.

In the preliminary portion of the study, therefore, the task of analyzing critically the mass of literature on reading and study was undertaken for the purpose of obtaining specific

suggestions which might be incorporated into a set of check lists which would give a complete analysis of the reading and study activities of college students. As a result of this analysis a set of 19 check lists was formulated, containing 931 specific items stated in positive form, which summarized the best procedures in reading and study suggested by experts in some 400 books, articles, and investigations. A five-point scale of responses to the items was adopted to obtain a measure of the relative degree or quality of each response. The responses were obtained therefore in terms of never, seldom, often, usually, and always. Since the items were stated in positive form and were good procedures, the responses never and seldom indicated a major degree of deficiency on the item involved.

The analysis of previous studies also showed that many factors or abilities other than the actual reading and study habits of students were of considerable importance in conditioning scholastic success and reading and study achievement. In order to obtain a complete and comprehensive diagnosis, additional test and exploratory materials were adopted which attempted to measure the related areas that had diagnostic significance. These diagnostic materials consisted of the following: (1) the University placement tests in reading, English, personality, and intelligence or scholastic aptitude; (2) 21 informal individual tests of abilities intimately related to reading and study, such as vision, hearing, association, and memory; (3) personal and family data sheets containing 224 exploratory items, the responses to which gave an added understanding of the individual in general and of the areas of experience, training, or environment which condition or affect general reading and study achievement; (4) a series of informal ratings on scholastic items, personal traits and characteristics, and personal appearance and hygiene, which represented a composite of the ratings made on these traits by high-school principals and teachers, college advisers and instructors, and the investigator. These diverse types of diagnostic materials provide a relatively complete diagnosis of reading and study abilities.

These diagnostic materials were partially standardized by administering them to four groups of students: (1) typical samples of 100, carefully selected to represent the total population of 700 students of the class of 1932-33; (2) 50 superior

students; (3) 50 middle students; and (4) 50 inferior students. The scores derived from the administration of the instrument to the typical samples of 100 were used to construct a five-point comprehensive rating scale upon which the scores of any student might be projected, subsequently, and reduced to a five-point basis. This rating device enabled the investigator to combine or summate scores on a comparable basis without resorting to the use of complex standard scores. The scores derived from the administration of the diagnostic materials to the 50 superior, 50 middle, and 50 inferior students were used to construct a three-point classificatory scale upon which the scores of a student on any portion or all of the diagnostic materials could be projected and classified as superior, middle, or inferior. The middle group was used primarily to delimit the middle area of the scale, setting the upper limits for the inferior and the lower limits for the superior students. The use of this scale furnished a convenient means of identifying readily the inferior students who were in need of remediation in reading and study.

Significant differences were found between the performance of the superior and inferior students on all of the diagnostic materials, with the exception of the group of personal ratings. The final rating scale score on the diagnostic technique was 4.2 for the superior group and 1.9 for the inferior group, or a difference of $2.3 \pm .311$ on a five-point scale. Approximately half of the entire area of the scale lay between the mean scores of the two groups. This difference was more than 7 times its probable error. The comprehensive diagnostic technique is therefore a significantly reliable instrument for differentiating between superior and inferior students. Since these higher scores achieved by the superior students are, at the same time, representative of a higher degree of ability on the diagnostic materials, the instrument is, in addition, a valuable instrument for the diagnosis of individual deficiency.

The personal and family data sheets were not of a type that could be standardized. They were developed only for the purpose of exploring the background of training and environment of the individual, as a supplementary form of diagnosis which would give the remedial worker a more complete understanding of the student. Approximately 119 of the items were found to be of considerable significance, however, in differentiating between

superior and inferior students. These significant items were reported in detail in Chapter III. When any considerable number of favorable or unfavorable responses on these data sheets occur, concomitantly, with favorable or unfavorable scores on the standardized materials, they form a valuable additional means of diagnosis and differentiation.

The diagnostic instrument was designed to be used, primarily, as an informal, individual method of determining reading and study deficiencies, for purposes of subsequent remediation. When used for this purpose, the time element is not of major importance, since a remedial worker should be prepared to spend considerable time and effort on the diagnosis of each case. The administration of the diagnostic instrument, other than the placement tests, can be accomplished, individually and orally, in approximately two and one-half hours. The reading and study check lists can be administered as a group measure by simply placing the words "Do you" before the individual items, before each group of items, or at the beginning of the complete set of check lists. When used in this form, the lists can be handed to the students in printed form and administered in a manner similar to that used with other group measures. The latter method was not used by the investigator in the establishment of the norms and experimental criteria since it is subject to the common deficiencies and unreliability of the questionnaire method.¹

The Application of the Diagnostic Instrument and Related Remedial Procedures under Experimental Conditions to Measure Their Administrative Value

Numerous books on "How to Study" had been published before this investigation was begun. These books were written largely, however, on an a priori basis supplemented by the observations and experience of the authors. None of the investigators had formulated a comprehensive diagnostic instrument or subjected their remedial suggestions to a detailed and adequately controlled experimental investigation to measure their administrative value.

The problem considered in the final portion of the present study, therefore, was the application of the diagnostic instrument

¹Leonard V. Koos, The Questionnaire in Education, pp. 144-67. New York: The Macmillan Company, 1928.

and related remedial procedures to potential college failures under experimental conditions to measure their administrative value. This involved four major steps: (1) the preliminary identification of potential college failures; (2) the administration of the diagnostic instrument to identify the students who had reading and study deficiencies; (3) the setting up of adequate experimental controls; and (4) the introduction of the experimental variable, which, in this study, was the remediation of reading and study deficiencies.

The complete battery of diagnostic materials was too lengthy to be administered by the investigator to the entire class of 700 students entering the University in the autumn quarter, September, 1933. A group of 39 potential college failures was identified therefore by means of the statistical predictive methods described in Chapter IV. The application of an additional experimental qualification, that of registration in courses which would permit of the selection of matched pairs of students in the same college classes, eliminated 13 of the group of 39 subjects.

With the permission and cooperation of Dean Brumbaugh the remaining 26 students were summoned to the office of the Dean of Students, the purpose of the experiment was explained, and the diagnostic materials were administered during the two weeks beginning October 23 and ending November 4. A further classification of these students was made on the basis of their performance on the diagnostic materials and 3 students were eliminated because they did not have a notable degree of specific deficiency in reading and study. A final group of 23 students was identified who were willing to cooperate in the experiment and who had marked reading and study deficiencies.

The experiment was organized by dividing the subjects into two experimental groups and one control group, consisting of 10, 3, and 10 subjects, respectively. The matched pair and parallel group methods of experimentation were used. Each pair was matched on 10 factors, as follows: percentile rank on the American Council on Education Psychological Examination, average of high-school marks, rank in high-school graduating class, percentile rank on the Iowa Comprehension Test, letter mark on the Thurstone Personality Schedule, age, sex, place of residence, course registration, and performance on the complete diagnostic instrument in terms of the final rating scale score of each stu-

dent. The additional factors of registration in English 102 and supplementary advisory service were held constant, since all of the subjects were in the course and received the regular advisory service. The significant variable introduced into the experiment was the remedial training in reading and study given by the investigator.

After the organization of the experiment and the preliminary identification of the subjects, a detailed diagnosis was obtained for each experimental subject by an analysis of the specific responses made by them to the items on the tests and check lists which compose the diagnostic instrument. On the basis of this specific diagnosis the remedial program was organized and conducted.

No remedial training was given to the control subjects. A self-directed-study or self-teaching method was used with Experimental Group B. These students were given complete sets of the reading and study check lists on which the specific items of deficiency were indicated for each subject and the proper remedial suggestions were outlined. In addition, the students were given the vocabulary lists described in Chapter V as a means of improving their vocabularies.

Relatively intense, individual, or clinical remedial procedures were used with the 10 subjects in Experimental Group A, for a period of 29 weeks. Each subject participated with the investigator in a one-hour remedial conference each week. The reading and study check lists and the basic vocabulary lists were used as the major remedial materials. As shown in Chapter V, the check lists constituted a complete guide to the best procedures in reading and study, as well as a means of identifying the specific items of deficiency. The vocabulary lists were used by the students, themselves, as means of improving their vocabularies. Supplementary practice drill and exercises of various types were given when needed and the subjects were informed concerning the use of special remedial devices.

At the close of the diagnostic and remedial experiment, a final testing program was administered to the experimental and control subjects to obtain a measure of their improvement during the training period. Two criteria were used to measure the value of the experimental program: (1) improvement in reading and study as measured by the specific tests used; and (2) the performance

of the subjects on the comprehensive course examinations administered by the Board of Examinations. Comparisons were drawn between the performance of the three groups on these measures. Since a reliable prediction of F or failure had been made for all the subjects, the administrative value of the experimental program was measured, primarily, in terms of the marks actually earned on the comprehensive examinations.

It was found that Experimental Group A did significantly better than the Experimental Group B or the Control Group on all of the experimental criteria. Similarly, the Experimental Group was significantly better than the Control Group. The mean per cent of improvement in reading and study made by the three groups during the experiment was 54.3 for the Experimental Group A, 44.2 for the Experimental Group B, and 5.8 for the Control Group. The significance of these gains was evaluated further by projecting the scores and gains upon the comprehensive five-point rating scale. It was found that the mean gain for Experimental Group A on this scale was 2.1, or approximately half of the entire area of the scale. The scale gain of Experimental Group B was 1.1, and of the Control Group, 0.2.

Since the Control Group was matched with the Experimental Groups on all of the essential variables, as shown in Chapter IV, the gain made by the Control Group was interpreted as a measure of the chance variation. The scale gains made by the Experimental Groups A and B were 10.5 and 5.5 times, respectively, the chance variation or gain made by the Control Group. The experimental program thus possessed considerable value in securing improvement in reading and study habits.

The individual or clinical method of remediation was consistently and significantly superior in its results to the self-directed-study or self-teaching procedure. The final rating scale difference in favor of Experimental Group A was 1.0, or 5 times the chance variation. Since the number of subjects in the self-directed-study group was too small for adequate statistical treatment, no general conclusion was drawn from this aspect of the experiment. The self-directed-study procedure was found to possess considerable value, however, as shown by the experimental results, and may be preferred in administrative situations where financial stringencies or the inadequacy of the staff demand the

use of such procedures. The measures developed in the study may be used in either type of program.

The actual administrative value of the diagnostic instrument and related remedial procedures was measured in terms of the percentage of F or failing marks actually earned on the comprehensive examinations by the three groups, as compared with the statistical prediction of F, which had been made previously, and with the percentage of F marks earned by similarly selected groups of the two previous "New Plan" classes, 1931-32 and 1932-33.

It was found that the percentage of F or failing marks, earned by the three groups used in the experiment, was 13.3 for Experimental Group A, 44.5 for the Experimental Group B, and 73.3 for the Control Group. The predicted percentage of F marks for these groups, as shown by the percentage of F's earned by the two similar groups of the previous classes, was 78.2. The Control Group did slightly better than the prediction. The difference between the percentage of F marks predicted and actually earned by the three groups was 64.9, 33.7, and 4.9 for Experimental Group A, Experimental Group B, and the Control Group, respectively. The 4.9 per cent gain made by the Control Group was interpreted as due to chance variation, a slight error in the predictive methods, or to the suggestions received during the initial and final testing programs. On the basis of this interpretation, the scholastic gains made by Experimental Groups A and B were 13.2 and 6.9 times, respectively, the chance variation. It is apparent, therefore, that there is considerable administrative value in the materials developed in the study.

The Contributions and Limitations of the Study

The general need for a comprehensive diagnostic instrument in reading and study, at the college level, has already been emphasized. The major contribution of the present study consists in the development of such an instrument, its partial standardization, and the experimental determination of the administrative value of the diagnostic instrument and related remedial procedures.

The contribution of the study lies not so much in the development of new diagnostic and remedial suggestions or in the discovery of new relationships between reading and study abilities and scholastic success. A wealth of material of this nature was at hand in the literature on these abilities. The general con-

tribution of the study rests in the fact that the diagnostic materials and suggestions which had been presented in a multitude of previous studies were brought together into an integrated and comprehensive diagnostic instrument. However, these materials were subjected to a further empirical investigation to determine their actual administrative value.

The comprehensive nature of the instrument and of the diagnosis which may be obtained through its use may be seen readily from the description of the diagnostic materials in Chapter III. These materials give both a general and specific diagnosis or analysis of reading and study abilities, as well as of the related factors of experience, training, environment, and personality which previous studies have shown to be of significance in conditioning scholastic success and reading and study achievement. Since the effectiveness of subsequent remedial procedures depends very largely upon the adequacy of the diagnosis, the value of the instrument for this purpose is readily apparent.

The present investigation is the first in which comparative norms were developed for the diagnostic materials. A Comprehensive Rating Scale was constructed by means of which the scores of a student on any of the diagnostic materials, with the exception of the data sheets, may be reduced to a comparable five-point basis. A Classificatory Rating Scale was constructed, also, by means of which students may be definitely classified as superior or inferior. The use of these two statistical devices enables an administrative officer to identify inferior students readily for purposes of subsequent remediation.

Previous investigators had not subjected their diagnostic and remedial procedures to adequately controlled experimental conditions to measure their administrative value. This was done in the present study and it was found that the experimental program was of significant value both as a means of improving the students' personal efficiency in reading and study and their scholastic success on the college comprehensive examinations.

Other facts which add to the significance of the contribution made in the present study may be indicated by the following facts: (1) it is the first comprehensive investigation of its type to be made under a "New Plan" type of college organization; (2) the success of the experiment is measured by an exceedingly rigid type of criterion, that of success on the new-type compre-

hensive examinations; (3) it is the first study in which potential rather than actually failing students were used; and (4) it is the first study in which a considerable portion of the diagnostic materials were used, in turn, as remedial materials. This last factor is of considerable importance, since a diagnosis is of little value if the appropriate remedial suggestions or procedures are not available. The reading and study check lists developed in the study give not only a complete and specific diagnosis of the reading and study habits or activities of students, but furnish, in addition, a complete set of remedial guides to the best procedures. They furnish therefore a convenient basis for the organization of a remedial program.

Many recent innovations in higher institutions reflect the growing tendency on the part of college administrators to adjust the individual student to the institution. The American philosophy of mass education has been of prime importance in bringing about such innovations and present economic and social trends¹ point to an expansion in the application of this principle. Similarly, the fundamental postulates of the "New Plan" of the University of Chicago recognize the existence of individual differences and attempt to make some adjustment in the light of these differences. The major contribution and educational significance of the present study lies, therefore, in the fact that it develops and partially standardizes a comprehensive diagnostic instrument and related remedial procedures which may be used as actual administrative means of adjustment or guidance in two fundamental types of academic activity. Through the use of the experimental program presented in this study, those individual differences which are due primarily to the effective use of good reading and study habits, activities, or procedures, may be lessened to a considerable degree. The inferior students may be given aid which will enable them to compete more effectively with the average or superior students.

As shown in Chapter V, the experimental program was of significant value in promoting the retention of potential college failures. Since the average mortality due to scholastic failure

¹Charles H. Judd, Problems of Education in the United States. Recent Social Trends Monograph. New York: McGraw-Hill Book Company, Inc., 1933.

under the "New Plan" is more than one-fifth of all entering freshmen, the value of the materials developed in the study is apparent. If absolute standards of performance are used in the marking of the college examinations, rather than the relative standards¹ used in the past, the retention problem will be lessened to a considerable extent through the use of the procedures used in the experimental portion of this study. The materials developed in the study can be used by the regular counselors or by special remedial workers, if such are available, as the basis for the program of diagnosis and remediation.

In connection with the educational significance of the study, however, certain limitations must be noted. One of the major limitations of the diagnostic instrument, in the opinion of some college administrators, is the time necessary for its administration, which is approximately three and one-half hours. This time limitation is not as important as it seems at first thought, however, since norms are available for each portion of the diagnostic materials as well as for the entire battery of materials. Since the adequacy of a remedial program depends to a considerable degree upon the adequacy of the diagnosis, it is recommended that the entire battery be used when possible. The adjustment of the individual student rather than the time element involved should be the major consideration of a remedial worker.

It has been shown in Chapter III that the student body of the University of Chicago is distinctly superior to the student body of many other institutions. This fact introduces a further limitation upon the use of the diagnostic instrument. The norms derived at this institution will be strictly applicable only in institutions which are on a similar intellectual level. This difficulty is readily overcome, however, if institutions having a relatively lower quality of students desire to use the diagnostic instrument. The complete method of deriving norms has been described in Chapters II and III of the present study. The diagnostic materials can be administered therefore in a manner similar to that used in this study, and the appropriate norms can be established readily for the particular institution.

Other limitations of the study which should be noted are concerned primarily with the scope of the study rather than with

¹Reeves and others, op. cit., V, 26.

the use of the diagnostic instrument. These are: (1) the study does not show what results can be achieved through the application of the experimental program to average or superior students who have good general ability but who possess some specific deficiencies which might be discovered through the use of the diagnostic materials; (2) the classificatory or differential function of the rating scales is subordinated to the diagnostic function; (3) the study does not show what results might be achieved through the use of group remedial training as compared with the individual or clinical remedial procedure; and (4) the small number of subjects in Experimental Group B, the self-directed-study group, make the results of this type of remedial training statistically inconclusive.

The present study was so comprehensive, however, in its present form that the addition of the foregoing considerations might have complicated the procedure seriously. The use of larger groups would have reduced the adequacy of the experimental controls to a very considerable degree, since it was difficult to equate the experimental and control subjects on the essential variables when the small groups were used. These limitations in the scope of the study are considered therefore in the final section of the chapter in which suggestions are given for the use and further study of the diagnostic materials.

Suggestions for the Use and Further Study of the Diagnostic Materials

There are two potential uses to which the diagnostic materials developed in the study may be put. They may be used to identify and give assistance to those students who are likely to be retarded in college or who are potential college failures. They may be used, in addition, to identify potentially superior students in order to increase their personal efficiency or to accelerate their progress.

The major emphasis of the present study was placed upon the remediation and retention of potential college failures. There is some experimental evidence in previous investigations¹

¹J. S. Butterweck, "The Problem of Teaching High-School Pupils How to Study." Unpublished Doctor's Dissertation, Department of Philosophy, Columbia University, 1926.

to show that potentially superior students may achieve very superior results through the use of diagnostic and remedial procedures in reading and study. The materials constructed in the study may be used in either case as the necessary basis of diagnosis, differentiation, and remediation.

On the basis of the experience and evidence acquired in this investigation, it is recommended that the diagnostic materials be used as follows:

1. The regular program of placement tests should be administered by the proper administrative officers. The placement tests included in the diagnostic instrument are the American Council on Education Psychological Examination, the Iowa Reading Test, the Minnesota Reading Examination, the Thurstone Personality Schedule, and the University of Chicago English Placement and Qualifying Examination, 1932 edition.

2. The reading and study check lists and the personal and family data sheets may be administered in any one of several ways: they may be administered with the regular placement tests as student check lists or student questionnaires; the college advisers may administer them during the first conference with each student following registration; the advisers may administer them by handing the lists to the students at the time of the first conference and requesting that the lists be checked properly and handed to the adviser at the next conference; they may be administered over a long period by making each list the basis for an individual diagnostic and remedial conference; and they may be used as the basis for diagnosis and remediation during a formal "How to Study" course, if such a procedure is desired in preference to an individual counseling program.

3. When the desired portion of the diagnostic materials has been administered, the scores may be tabulated, projected and rated on the comprehensive rating scale, and classified into the proper categories on the classificatory scale. The students who have general deficiencies in reading and study may be identified readily by this procedure, since they will be classified as inferior on the diagnostic materials in which a major degree of deficiency is present.

4. When the students who have general reading and study deficiencies have been identified for purposes of subsequent remediation, a further and more intensive diagnosis should be made

to determine the specific items of deficiency. This may be accomplished by analyzing the specific responses of the students to the actual items of the diagnostic materials, as illustrated in Chapter V. When the specific items of deficiency are located, they will become the basis upon which the subsequent remedial procedures may be instituted. The reading and study check lists may be used as the major source of remedial suggestions and as a comprehensive guide to the best reading and study procedures. The lists may be used as self-directed-study guides by handing them to the students, indicating the specific items of deficiency, and suggesting the proper remedial procedures, or they may be used as the basis for a series of remedial conferences similar to those presented in Chapter VI. When the diagnostic and remedial program is conducted at the University of Chicago it is recommended especially that the basic vocabulary lists developed in the study be used to remedy vocabulary deficiencies as an indirect means of improving reading ability.

The foregoing procedure summarizes the general use to which the materials developed in the study may be put. Since norms are available for the complete battery of diagnostic materials and for each subdivision, the diagnostic and remedial program may be as limited or as comprehensive as is desired in the particular administrative situation.

As the present study progressed it was found that the differential or classificatory value of the diagnostic instrument had to be sacrificed, in part, in order to retain the comprehensive quality of the diagnosis. Many of the items of the tests and check lists were found to be of less differential value than other items. If it is desired to obtain a more sharply defined differentiation between superior and inferior students, the scores made by these two groups on all items of the tests and check lists may be analyzed and the items which prove to be of less differential value may be eliminated from the instrument. Summations of the scores made by the superior and inferior students on the diagnostic materials may be computed, finally, on the new basis and projected upon the rating scales to obtain new normative values.

The items or tests of highest differential value may be combined further into a formal standardized test to be used purely as a means of differentiation or classification and not as the

basis for a comprehensive diagnosis. The results achieved by this elimination of items or tests of least differential value will enhance the classificatory value of the instrument to a considerable degree, but will decrease the diagnostic value proportionately.

A further extension of the present study may be made by organizing and conducting a controlled diagnostic and remedial experiment with average or superior students instead of potential college failures. Methods of prediction and remediation similar to those employed in the present study may be used. An experiment of this type would give objective data for use in evaluating the amount of student acceleration that could be achieved under "New Plan" conditions through the use of diagnostic and remedial procedures in reading and study with average or superior students.

The administrative value of the diagnostic instrument and related remedial procedures may be measured further by organizing an experiment that is identical with that of the present study, with the exception that a group remedial procedure be used instead of the clinical or self-directed-study methods. Such an experiment would give a critical evaluation of the use of the materials developed in the present study as a basis for group remediation, which is the form in which many college administrators may desire to administer such a program. The remedial measures probably could be conducted on a more economical basis, financially, through a group procedure than through the individual methods used in the present study. The question of the comparability of results which might be obtained from the group as contrasted with the individual methods would be answered through an adequate experiment of the type suggested.

The limitations imposed by the small number of subjects that was used in the Experimental Group B, or self-directed-study group, made it impractical to draw any thoroughly justified general conclusions concerning this phase of the experiment. A further experiment, using adequate controls and a larger group of subjects, would be of considerable value in determining more adequately the value of the diagnostic instrument and related remedial procedures in a self-directed-study or self-teaching situation. This extension of the study is important, since many administrators may wish to use the materials developed in the study in this manner, owing to the greater cost of the individual or clinical remedial program.

